

1. Relacja między danymi a informacją.

Dane odpowiadają pojedynczym zarejestrowanym faktom dotyczącym zjawisk w otaczającym nas świecie, a informacja to rozszerzenie wiedzy na podstawie danych. Informacje docierają do nas poprzez proces interpretacji danych i mogą nam pomóc zmienić lub też poprawić zrozumienie (czyli generalnie naszą wiedzę).

Wg definicji Laudon&Laudon

Dane to surowe fakty, które mogą być kształtowane i formowane by stworzyć informację. Informacja to dane które zostały uformowane przez człowieka w istotną i użyteczną postać.

2. Różnica między scentralizowanymi a rozproszonymi bazami

Bazy scentralizowane to takie w których wszystkie dane znajdują się w jednej fizycznej lokalizacji pod kontrolą jednego SZBD. W bazach rozproszonych natomiast dane mogą znajdować się w różnych fizycznych lokalizacjach i pod kontrolą różnych SZBD. W rozproszonej bazie danych zachodzi lokalna autonomia tzn. każdy fragment danych to odrębna baza danych.

3. Definicje partycjonowania i replikacji

Partycjonowanie to możliwość podzielenia pojedynczej bazy danych na fizycznie odrębne fragmenty w celu optymalizacji wydajności, poprzez zwiększenie efektywności dostępu do danych np. poprzez zmniejszenie rozmiaru danych które trzeba przeszukać lub alokowanie fragmentów "blisko" miejsca ich wykorzystania (redukcja kosztów transmisji sieciowej). Jest niewidoczne dla użytkownika i pozostaje pod kontrolą SZBD.

Replikacja to wymiana (poprzez kopiowanie) danych pomiędzy różnymi lokalizacjami np w celu przyspieszenia dostępu do danych poprzez umieszczenie danych w lokalizacji w której są intensywnie wykorzystywane (np. tabele słownikowe). Użytkownik może operować w taki sam sposób na danych oryginalnych (źródłowych) jak i na kopii. Podstawowym obiektem baz danych, który się replikuje jest tabela.

4. Fakty i wymiary w kontekście hurtowni danych, ich modele

Fakty opisują zdarzenia (np. sprzedaż) przeważnie w ujęciu ilościowym (np. ilość, wartość, cena), i znajdują się w głównej tabeli Hurtowni Danych połączonej z tabelami wymiarów. Wymiary nadają znaczenie faktom (np. kto, co, komu, kiedy) i przechowywane są w odrębnych tabelach. Mogą mieć strukturę hierarchiczną. Ich relacje zwykle można przedstawić w jednym z 3 schematów: gwiazdy, płotka śniegu (znormalizowane tabele wymiarów), konstelacji (w wypadku istnienia więcej niż jednej tabeli faktów, może być wykorzystany zarówno w schemacie gwiazdy, jak i płotka śniegu).

Modele Hurtowni Danych to ROLAP- relacyjna- dane oraz tabele wymiarów przechowywane są w relacyjnych bazach danych MOLAP- wielowarstwowa- dane są organizowane w postaci wielowymiarowych kostek, mają dużo większą wydajność niż ROLAP, posiadają istotną wadę- możliwość przechowywania znacznie mniejszej ilości danych niż w ROLAP HOLAP- hybrydowa