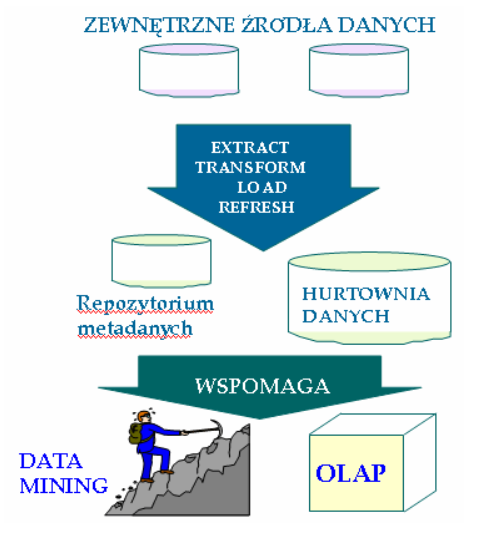
20. Proces i architektura hurtowni danych, wielowymiarowy model danych.

**Hurtownia danych** to swojego rodzaju baza danych, mająca służyć wspomaganiu procesu podejmowania decyzji, która jest:

* Zorientowana na temat – oznacza to, że zbierane dane dotyczą konkretnego tematu np. sprzedaży, a nie działań np. zbierania zamówień.
* Nieulotna – oznacza to, że dane raz umieszczone w hurtowni, zazwyczaj pozostają niezmienione. Każdy użytkownik bazy danych ma pewność, że zapytanie zawsze zwróci taki sam wynik, niezależnie od tego jak często są wykonywane.
* Zintegrowana – oznacza to, że dane są jednolite, np. daty przechowywane są zawsze w tym samym formacie, znaki kodowane są w ten sam sposób.
* Zróżnicowana czasowo – oznacza to, że gromadzone są dane historyczne. Prawie wszystkie zapytania kierowane do hurtowni danych wymagają prześledzenia jakiegoś odcinka czasu.

**Architektura hurtowni danych** – hurtownia danych pobiera dane z zewnętrznych źródeł danych. Zanim dane znajdą się w hurtowni muszą zostać pobrane, ulec transformacji, zostać zintegrowane i odwzorowane, a w końcu załadowane do hurtowni. Poza ładowaniem początkowym, dokonywane są ładowania przyrostowe w celu odświeżania hurtowni. Do ekstrakcji, transformacji i ładowania służy narzędzie ETL, które może być częścią hurtowni bądź też narzędziem zewnętrznym. Należy zdawać sobie sprawę z tego, że dane do hurtowni mogą być pobierane z wielu różnych źródeł.

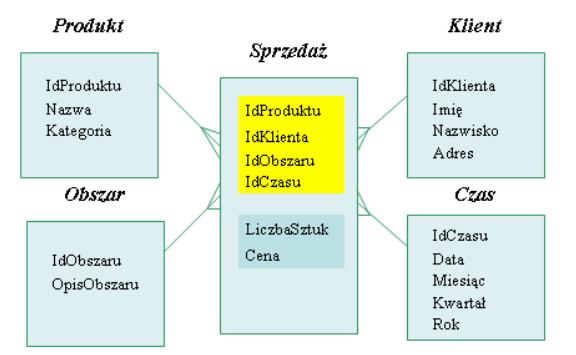
Hurtowni danych towarzyszy mocno rozbudowane repozytorium metadanych, w którym znajdują się wszystkie informacje z poziomu pojęciowego (opis zawartości, opis procesu integracji pojęciowej itp.), logicznego (nazwy tabel i kolumn, typ danych, rodzaje związków między tabelami itd.) i fizycznego (definicja fizycznego rozmieszczania danych i ich przepływu, informacja o lokalizacji fragmentów), statystyki danych i uzycia oraz informacje administracyjne (zasady dostępu do danych, definicje użytkowników i ich grup, terminy wykonywania cyklicznych czynności).

**Wielowymiarowy model danych**

**Schemat gwiazdy** – pozwala przedstawić dane w postaci wielowymiarowego modelu danych.

Centrum schematu gwiazdy tworzy **tabela faktów**, czyli tabela gromadząca informacje o temacie wokół którego skupia się hurtowania. Faktów dotyczy większość zapytań zadawanych w hurtowni danych. Tabela faktów jest znormalizowana, jej klucz główny stanowią klucze obce pochodzące z pozostałych tabel, natomiast kolumny niekluczowe, czyli tzw. **miary** muszą być sumowalne. Patrząc od strony biznesowej w tabeli faktów gromadzone są informacje dotyczące tego, co firmie przynosi zysk. Miary to w pewnym sensie wyznaczniki tego zysku, czyli liczba sprzedanych sztuk, cena.

Tabelę faktów otaczają **tabele wymiarów**, zawierające dane o wymiarach analizy, czyli o tym wg czego będziemy analizowali dane z tabeli faktów. Cechą charakterystyczną tabel wymiarów jest brak normalizacji. Np. w tabeli produkt znajdują się wszystkie potrzebne z punktu widzenia przyszłej analizy informacje o produkcie, łącznie z informacjami o kategorii produktu. Szczególnym wymiarem analizy jest wymiar *czasu*. Wymiary często mają strukturę hierarchiczną np. produkt-kategoria, data-miesiąc-kwartał-rok, klient-miasto-kraj.



Istnieją jeszcze inne modele jak:

| **Schemat płatka śniegu** | **Schemat konstelacyjny**  **Obraz zawierający tekst, koperta, wizytówka  Opis wygenerowany automatycznie** |
| --- | --- |