



Hurtownie danych - przegląd technologii

Robert Wrembel
Politechnika Poznańska
Instytut Informatyki
Robert.Wrembel@cs.put.poznan.pl
www.cs.put.poznan.pl/rwrembel



Metadane

- ⇒ **Biznesowe**
 - słowniki pojęć biznesowych
 - odwzorowania pojęć biznesowych w obiekty HD
 - reguły biznesowe
 - jakość danych
- ⇒ **Techniczne**
 - schematy
 - struktury fizyczne
 - charakterystyki danych
 - organizacja przestrzeni dyskowej
 - implementacje algorytmów (transformacje, czyszczenie)



Metadane

- ⇒ **Sterujące wykonaniem ETL**
 - harmonogramy
 - skrypty
 - logi z wykonania
 - monitorowanie
- ⇒ **Opisujące zewnętrzne źródła danych**
 - rodzaj źródła (relacyjna bd, obiektowa bd, xml, html, arkusz kalkulacyjny, ...)
 - struktura/schemat
 - metody dostępu
 - użytkownicy i prawa dostępu
 - wyniki analizy źródeł
 - dzienny przyrost danych
 - rozmiary danych



Metadane

- ⇒ **Opisujące procesy ETL**
 - struktura przepływu pracy
 - odwzorowania źródło ⇔ HD
 - odwzorowania rekordów źródłowych w docelowe (lineage)
 - definicje transformacji (nazwa, realizowany cel, wejście, wyjście, algorytm)
 - skrypty i zadania (nazwa, realizowany cel, źródło, struktury docelowe, pliki logów, pliki sterujące, statystyki efektywnościowe z wykonania, obsługa wyjątków/awarii)
 - harmonogram uruchamiania ETL (częstotliwość, obsługa wyjątków/awarii, pliki logów, statystyki efektywnościowe z wykonania)
 - logi z pracy ETL
 - charakterystyka danych
 - fizyczna organizacja przestrzeni dyskowej



Metadane

- ⇒ **Opisujące hurtownię danych**
 - **schemat (wymiary, hierarchie, fakty)**
 - **użytkownicy i uprawnienia**
 - **charakterystyka danych (rozkłady, histogramy, rozmiary)**
 - **struktury fizyczne (indeksy, klastry, perspektywy zmaterializowane/predefiniowane zapytania, partycjonowanie)**
 - **fizyczna organizacja przestrzeni dyskowej**



Metadane

- ⇒ **Opisujące użytkowników**
 - **profile użytkowników**
 - **subskrypcje raportów**
 - **definicje pojęć biznesowych**
 - **preferencje wyświetlania raportów**
 - **ustawienia portletów/kokpitów menadżerskich**



Standardy opisu metadanych

- ⇒ **Open Information Model (OIM)**
 - rozwijany przez Metadata Coalition
 - wspierany przez Microsoft, Brio Technologies, Informatica, SAS Institute
- ⇒ **Common Warehouse Metamodel**
 - rozwijany przez Object Management Group
 - wspierany przez IBM, Oracle, Hyperion
- ⇒ **W 2000 OIM zintegrowany z CWM ⇒ oba rozwijane jako CWM**
 - Common Warehouse Metamodel (CWM) Specification, 2001



CWM

- ⇒ **CWM podstawy**
 - zbiór pakietów metamodeli wspólnych dla pozostałych
- ⇒ **Relacyjne źródła danych**
 - opis relacyjnych źródeł danych i źródeł dostępnych przez ODBC/JDBC
- ⇒ **Rekordowe źródła danych**
 - opis struktur rekordowych (w bazach danych, plikach, językach programowania)
- ⇒ **Wielowymiarowe źródła danych**
 - opis struktur wielowymiarowego modelu danych (wymiar, hierarchie, kostki)
- ⇒ **Źródła danych XML**
 - opis źródeł danych XML
- ⇒ **Transformacje danych**
 - opis typowych transformacji struktur i danych pomiędzy modelami (relacyjny, obiektowy, XML, wielowymiarowy)



CWM

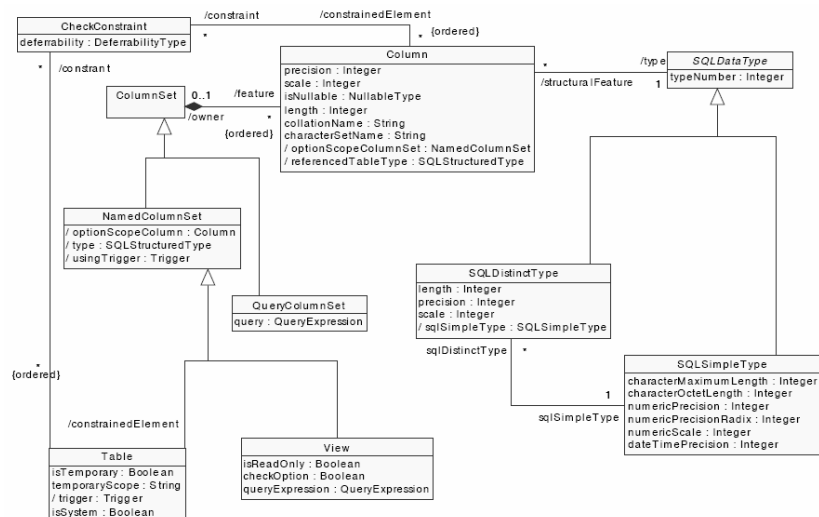
- ⇒ **OLAP**
 - definicje technologii, opis składowania danych, semantyka operacji roll-up, drill-down
- ⇒ **Eksploracja danych**
 - opis modeli eksploracji, wykorzystywanych atrybutów
- ⇒ **Wizualizacja danych**
 - opis technik wizualizacji danych
- ⇒ **Definicje biznesowe**
 - definicje pojęć biznesowych i ich odwzorowanie w struktury i dane
- ⇒ **Procesy**
 - opis procesów i zależności między nimi
- ⇒ **Praca HD**
 - opis struktur rejestrujących bieżącą pracą HD (statystyki, logi pracy procesów)

Robert Wrembel

9/13



CWM - atrybut

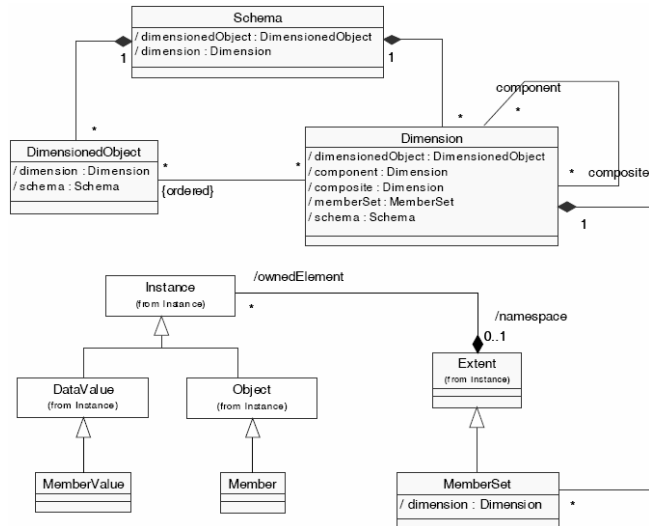


Robert Wrembel

10/13



CWM - wymiar

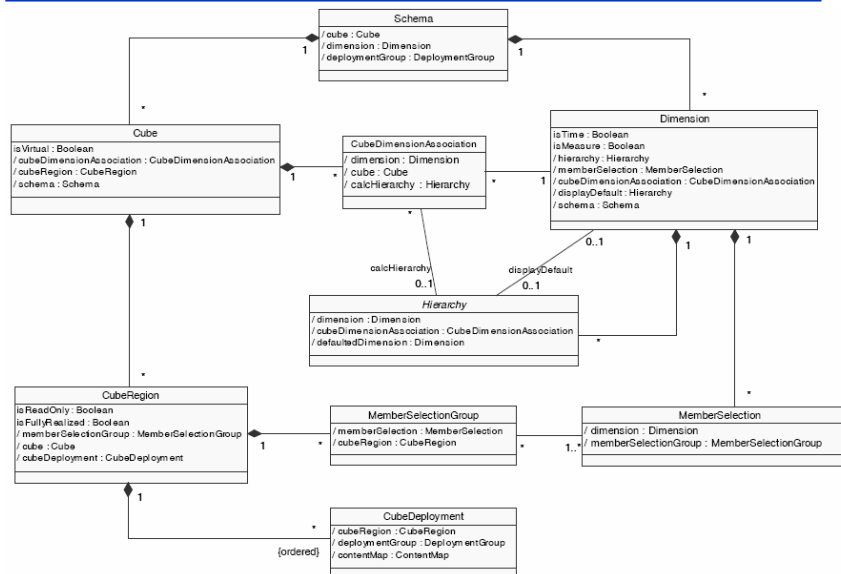


Robert Wrembel

11/13



CWM - kostka





CWM - hierarchia

