

Ewolucja technik modelowania hurtowni danych

Porównanie technik modelowania hurtowni danych – podsumowanie:

Strona 1/6		Podejście		
Kryterium	Podkryterium	Inmon	Kimball	Linstedt
Podstawowe informacje	Nazwa podejścia	Corporate Information Factory	Kimball Bus Architecture	Data Vault
	Typ podejścia	Z góry na dół	Z dołu do góry	Z góry na dół
	Normalizacja danych	Znormalizowane, trzecia postać normalna (3PN)	Dane zdenormalizowane	Znormalizowane, trzecia postać normalna (3PN)
	Redundancja	Brak	Tak	Brak
	Autor	William Harvey (Bill) Inmon (1945)	Ralph Kimball (1944)	Daniel E. Linstedt
	Ramy czasowe	W latach 70-tych Inmon rozpoczyna formułowanie definicji hurtowni danych. Podejście wypracowane w latach 90-tych ubiegłego stulecia.	Podejście wypracowane w latach 90-tych ubiegłego stulecia.	Pierwotnie zaprojektowane już w 1990 roku. Data Vault 1.0 wydane w 2000 roku jako metodologia modelowania. Data Vault 2.0 wydane w 2013 roku.

Strona 2/6		Podejście		
Kryterium	Podkryterium	Inmon	Kimball	Linstedt
Filozofia	Rozumienie hurtowni danych	Hurtownia to tematyczny, zintegrowany, zależny od czasu, trwały zbiór danych, ukierunkowany na wspomaganie procesów podejmowania decyzji. Hurtownia przechowuje dane „atomowe” na najniższym poziomie szczegółowości. Data marta tworzone są dopiero po utworzeniu pełnej hurtowni danych.	Hurtownia stanowi kopię danych transakcyjnych zapisanych w strukturach ukierunkowanych na zapytania analityczne. Mając na uwadze najważniejsze aspekty biznesowe lub działy, data marta są tworzone w pierwszej kolejności. Zapewniają one wąski widok danych organizacyjnych i, gdy jest to wymagane, można je połączyć w większą hurtownię danych.	Hurtownia jest zaprojektowana tak, aby zapewnić przechowywanie danych historycznych z wielorakich systemów źródłowych. Data Vault oznacza również, obok aspektu modelowania, sposób patrzenia na dane historyczne, który zapewnia audytowalność, śledzenie danych, szybkość ładowania oraz odporność na zmiany biznesowe.
	Główni aktorzy	Specjaliści IT	Użytkownicy końcowi	Użytkownicy końcowi
	Cel	Dostarczenie kompletnego rozwiązania technicznego opartego na udowodnionych metodach i technologiach.	Dostarczenie rozwiązania, które ułatwia bezpośrednie odpytywanie danych przez użytkowników końcowych.	Dostarczenie kompletnego rozwiązania opartego na udowodnionych metodach.

Strona 3/6		Podejście		
Kryterium	Podkryterium	Inmon	Kimball	Linstedt
Metodologia i architektura	Obszar przejściowy	Tak	Tak	Tak
	Architektura	Hurtowniane dane na najniższym poziomie granulacji są źródłem dla wydziałowych data martów.	Zestaw data martów tworzy hurtownię danych.	Data Vault dostarcza dane dla data martów.
	Układ danych	Dane relacyjne (3PN)	Dane relacyjne (płatek śniegu, gwiazda lub konstelacja faktów)	Dane relacyjne (3PN)
	Elementy charakterystyczne modelu	Brak	Wymiar, tabela faktów	Hub, link, satelita
	Złożoność techniki	Dość złożona	Prosta	Prosta
	Metodologia wytwarzania	Inspirowane przez model spiralny	Czterokrokowy proces: 1. Wybranie procesu biznesowego 2. Określenie granulacji danych 3. Identyfikacja wymiarów 4. Identyfikacja faktów	Oparte na metodyce zwinnej
	Koszty wdrożenia	Koszty początkowe wyższe, dalsze koszty niższe	Początkowe koszty niższe, każdy kolejny krok kosztuje tyle samo	Koszty projektu niższe niż w dwóch poprzednich podejściach
	Czas wytworzenia	Długi	Krótki	Krótki

Strona 4/6		Podejście		
Kryterium	Podkryterium	Inmon	Kimball	Linstedt
Integracja danych, procesy ETL	Integracja danych pochodzących z wielu źródeł	Transformacje muszą zostać zaplanowane w procesie ETL.	Transformacje muszą zostać zaplanowane w procesie ETL.	Separacja satelitów i kluczy biznesowych redukuje złożoność.
	Złożoność procesów ETL	Transformacje proste, jeśli model danych jest podobny do modelu źródeł danych.	Transformacje pomiędzy modelem OLTP a modelem wielowymiarowym są złożone.	Transformacje są proste do załadowania hubów, linków i satelitów.
	Orientacja	Oparte na danych	Oparte na procesach	Oparte na danych i procesach
	Narzędzia	Klasyczne narzędzia modelowania (ERD, DIS)	Modelowanie wymiarowe	Modelowanie hubów, linków i satelitów
	Zaangażowanie użytkowników końcowych	Słabe	Mocne	Mocne

Strona 5/6		Podejście		
Kryterium	Podkryterium	Inmon	Kimball	Linstedt
Zarządzanie cyklem życia	Elastyczność na zmiany w źródle danych	Potrzeba zmian w tabelach.	Częste zmiany w modelu źródłowym oddziałują na model hurtowni danych.	Zmiany w źródle nie wpływają na model hurtowni. Jedyna możliwa zmiana to dodanie odpowiedniej satelity.
	Elastyczność na nowe wymagania biznesowe	Model zmienia się tylko gdy potrzebne dane nie istnieją w hurtowni danych.	Nowe wymagania mają wpływ na model danych.	Brak zmian w modelu, wymagane dostarczenie danych do data martów.
	Łatwość zmiany modelu	Dane historyczne muszą być zmigrowane w niektórych przypadkach.	Refactoring niektórych tabel jest potrzebny w pewnych przypadkach.	Jedyna zmiana to dodanie odpowiednich satelit.
	Audytowalność	Dane historyczne są przechwytywane poprzez wstawianie nowych rekordów (dotyczy każdej zmiany).	Wykorzystywanie koncepcji wymiarów wolnozmiennych (ang. SCD = Slowly Changing Dimension) w celu zapisania zmian historycznych.	Dane historyczne mogą być przechwycone poprzez dodanie odpowiednich satelit. Natywna obsługa pełnej historyzacji danych.
	Wydajność zapytań	Słaba wydajność ze względu na strukturę w 3PN.	Model zaprojektowany w celu optymalizacji zapytań.	Bardzo wysoka standaryzacja jest powodem słabej wydajności. Wymagane marty w celach analizy i raportowania.

Strona 6/6		Podejście		
Kryterium	Podkryterium	Inmon	Kimball	Linstedt
Wykorzystanie modelu	Dostosowanie do prezentacji danych	Nie dostarcza warstwy końcowej użytkownika.	Schemat gwiazdy/ płotka śniegu idealny pod raportowanie.	Nie dostarcza warstwy prezentacyjnej dla użytkownika. Idealny do eksploracji danych, wymiary mogą być wygenerowane automatycznie.
	Odpytywanie modelu	Optymalna liczba złączeń dla wsparcia oszczędnego przechowywania (mniejsza redundancja).	Mało złączeń, trudne do rozwiązania problemy z różnymi poziomami ziarnistości danych.	Potrzeba bardzo wielu złączeń.

Autor:



Piotr Ziuziański

Senior Business Intelligence Specialist, JCommerce

Certyfikowany specjalista technologii Business Intelligence firmy Microsoft. Tworzy rozwiązania dla klientów, począwszy od etapu modelowania hurtowni danych, projektowania i wdrażania procesów ETL, po implementację modelu danych i wdrożenie wizualnej warstwy raportowej. Autor [bloga](#) poświęconego tematyce rozwiązań klasy Business Intelligence.