

Hurtownie Danych

Slowly Changing Dimension

Krzysztof Jankiewicz

Politechnika Poznańska

Instytut Informatyki

Slowly Changing Dimension

- Z założenia hurtownia danych zawiera dane trwałe – nieulotne.
- Nieulotność wymusza
 - wprowadzanie nowych danych do tabel faktów
 - uwzględnienie zmian wartości w atrybutach wymiarów
- Zazwyczaj stosuje się następujące podejścia do tzw. wymiarów wolnozmiennych (ang. *Slowly Changing Dimension*)
 - Type 0 – brak jakichkolwiek działań
 - Type 1 – nadpisywanie
 - Type 2 – tworzenie nowych wersji danych
 - Type 3 – dodatkowe atrybuty przechowujące np. poprzednią wersję wartości wybranych atrybutów
 - Type 4 – przechowywanie poprzednich wersji danych w oddzielnych strukturach

Slowly Changing Dimension type 0

- Raz utworzone wartości wymiarów nie są modyfikowane
- Zmiany w źródłach danych dla wymiarów są ignorowane

Rows of step: Adres zmieniony (3 rows)

^ #	KL_KLIENT_ID	KL_KOD_POCZTOWY	KL_MIASTO	KL_ULICA	klient_id	miasto	adres	kod_pocztowy
1	285	26209	Snowshoe	1336 Benin City Drive	285	Los Angeles	98 Stara Zagora Boulevard	90001
2	342	65583	Waynesville	1293 Nam Dinh Way	342	Eckert	1192 Tongliao Street	81418
3	399	77850	Concord	734 Bchar Place	399	Vera	953 Hodeida Street	74082

poprzednie wartości atrybutów

obecne wartości atrybutów

stałe wartości
atrybutów wymiarów

285	26209	Snowshoe	1336 Benin City Drive
342	65583	Waynesville	1293 Nam Dinh Way
399	77850	Concord	734 Bchar Place

Slowly Changing Dimension type 1

- Modyfikacja wartości atrybutów wymiarów w źródłach powoduje nadpisanie wartości atrybutów w hurtowni danych

Rows of step: Adres zmieniony (3 rows)

^ #	KL_KLIENT_ID	KL_KOD_POCZTOWY	KL_MIASTO	KL_ULICA	klient_id	miasto	adres	kod_pocztowy
1	285	26209	Snowshoe	1336 Benin City Drive	285	Los Angeles	98 Stara Zagora Boulevard	90001
2	342	65583	Waynesville	1293 Nam Dinh Way	342	Eckert	1192 Tongliao Street	81418
3	399	77850	Concord	734 Bchar Place	399	Vera	953 Hodeida Street	74082

poprzednie wartości atrybutów

obecne wartości atrybutów

przed aktualizacją

285	26209	Snowshoe	1336 Benin City Drive
342	65583	Waynesville	1293 Nam Dinh Way
399	77850	Concord	734 Bchar Place

po aktualizacji

285	90001	Los Angeles	28 Stara Zagora Boulevard
342	81418	Eckert	1192 Tongliao Street
399	74082	Vera	953 Hodeida Street

Slowly Changing Dimension type 2

Rows of step: Adres zmieniony (3 rows)

^	#	KL_KLIENT_ID	KL_KOD_POCZTOWY	KL_MIASTO	KL_ULICA	klient_id	miasto	adres	kod_pocztowy
1		285	26209	Snowshoe	1336 Benin City Drive	285	Los Angeles	98 Stara Zagora Boulevard	90001
2		342	65583	Waynesville	1293 Nam Dinh Way	342	Eckert	1192 Tongliao Street	81418
3		399	77850	Concord	734 Bchar Place	399	Vera	953 Hodeida Street	74082

poprzednie wartości atrybutów

obecne wartości atrybutów

przed aktualizacją

KL_ID	KL_KOD	KL_MIASTO	KL_ULICA	KL_SID	KL_VER	KL_START	KL_STOP
285	26209	Snowshoe	1336 Benin City Drive	1	1	1900-01-01	2999-12-31
342	65583	Waynesville	1293 Nam Dinh Way	2	1	1900-01-01	2999-12-31
399	77850	Concord	734 Bchar Place	3	1	1900-01-01	2999-12-31

po aktualizacji

KL_ID	KL_KOD	KL_MIASTO	KL_ULICA	KL_SID	KL_VER	KL_START	KL_STOP
285	26209	Snowshoe	1336 Benin City Drive	1	1	1900-01-01	2013-07-23
342	65583	Waynesville	1293 Nam Dinh Way	2	1	1900-01-01	2013-07-23
399	77850	Concord	734 Bchar Place	3	1	1900-01-01	2013-07-23
285	90001	Los Angeles	28 Stara Zagora Boulevard	4	2	2013-07-23	2999-12-31
342	81418	Eckert	1192 Tongilao Street	5	2	2013-07-23	2999-12-31
399	74082	Vera	953 Hodeida Street	6	2	2013-07-23	2999-12-31

Slowly Changing Dimension type 3

Rows of step: Adres zmieniony (3 rows)

^	#	KL_KLIENT_ID	KL_KOD_POCZTOWY	KL_MIASTO	KL_ULICA	klient_id	miasto	adres	kod_pocztowy
	1	285	26209	Snowshoe	1336 Benin City Drive	285	Los Angeles	98 Stara Zagora Boulevard	90001
	2	342	65583	Waynesville	1293 Nam Dinh Way	342	Eckert	1192 Tongliao Street	81418
	3	399	77850	Concord	734 Bchar Place	399	Vera	953 Hodeida Street	74082

poprzednie wartości atrybutów

obecne wartości atrybutów

przed
aktualizacją

KL_ID	KL_KOD	KL_MIASTO	KL_ULICA	KL_KOD_OLD	KL_MIASTO_OLD	KL_ULICA_OLD
285	26209	Snowshoe	1336 Benin City Drive			
342	65583	Waynesville	1293 Nam Dinh Way			
399	77850	Concord	734 Bchar Place			

po
aktualizacji

KL_ID	KL_KOD	KL_MIASTO	KL_ULICA	KL_KOD_OLD	KL_MIASTO_OLD	KL_ULICA_OLD
285	90001	Los Angeles	28 Stara Zagora Boulevard	26209	Snowshoe	1336 Benin City Drive
342	81418	Eckert	1192 Tongilao Street	65583	Waynesville	1293 Nam Dinh Way
399	74082	Vera	953 Hodeida Street	77850	Concord	734 Bchar Place

Slowly Changing Dimension type 4

Rows of step: Adres zmieniony (3 rows)

^	#	KL_KLIENT_ID	KL_KOD_POCZTOWY	KL_MIASTO	KL_ULICA	klient_id	miasto	adres	kod_pocztowy
	1	285	26209	Snowshoe	1336 Benin City Drive	285	Los Angeles	98 Stara Zagora Boulevard	90001
	2	342	65583	Waynesville	1293 Nam Dinh Way	342	Eckert	1192 Tongliao Street	81418
	3	399	77850	Concord	734 Bchar Place	399	Vera	953 Hodeida Street	74082

poprzednie wartości atrybutów

obecne wartości atrybutów

KLIENCI

ADRESY_KLIENTOW

przed
aktualizacją

KL_ID	KL_IMIE	KL_NAZWISKO
285	Adam	Nowak
342	Piotr	Kowalski
399	Zofia	Zajęc

AD_SID	AD_KOD	AD_MIASTO	AD_ULICA
1	26209	Snowshoe	1336 Benin City Drive
2	65583	Waynesville	1293 Nam Dinh Way
3	77850	Concord	734 Bchar Place

po
aktualizacji

KL_ID	KL_IMIE	KL_NAZWISKO
285	Adam	Nowak
342	Piotr	Kowalski
399	Zofia	Zajęc

AD_SID	AD_KOD	AD_MIASTO	AD_ULICA
1	26209	Snowshoe	1336 Benin City Drive
2	65583	Waynesville	1293 Nam Dinh Way
3	77850	Concord	734 Bchar Place
4	90001	Los Angeles	28 Stara Zagora Boulevard
5	81418	Eckert	1192 Tongilao Street
6	74082	Vera	953 Hodeida Street

Slowly Changing Dimensions 1-4

wpływ na tabele faktów

- Różne CDC mają różny wpływ na tabelę faktów
 - Type 0 – brak zmian
 - Type 1 – brak zmian
 - Type 2 – tabela faktów posiada klucz obcy oparty na sztucznym kluczu głównym tabeli wymiarów (ang. *surrogate key*)
 - Type 3 – bez zmian
 - Type 4 – dodatkowa tabela wymiaru, ze sztucznym kluczem głównym, wymaga dodatkowego klucza obcego w tabeli faktów

Własności SCD 1-4 – podsumowanie

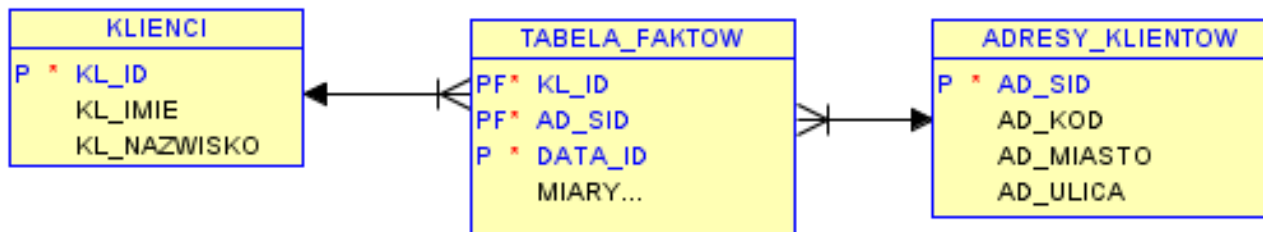
- Type 0 – (ignoruj)
 - ignorowanie zmieniających się wartości
- Type 1 – (nadpisz)
 - ignorowanie historycznych wartości
- Type 2 – (dodaj wiersz)
 - zapamiętywanie pełnej historii zmian
 - identyczne traktowanie atrybutów niezmiennych i tych, dla których zmiany zachodzą szybko
- Type 3 – (dodaj kolumnę)
 - częściowe uwzględnianie historii zmian (np. dwa ostatnie zestawy wartości, pierwszy i ostatni zestaw wartości)
- Type 4 – (dodaj mini wymiar)
 - zapamiętywanie pełnej historii zmian
 - rozróżnienie atrybutów niezmiennych i tych, dla których zmiany zachodzą szybko

Slowly Changing Dimension 5-7

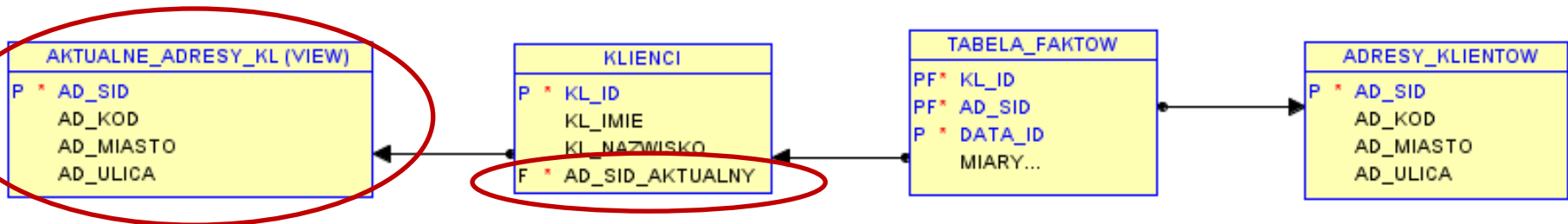
- Oprócz często wykorzystywanych SCD 1-4, stosowane są także bardziej wyrafinowane rozwiązania
 - Type 5 – dodatkowy mini wymiar + nadpisywany wskaźnik na bieżący zestaw wartości
 - Type 6 – nadpisywane atrybuty typu 1 + dodatkowe wiersze typu 2
 - Type 7 – typ 2 + dodatkowy wymiar zawierający bieżące wartości (type 1)

Slowly Changing Dimension type 5

- W SCD type 4 (przypomnienie):



- W SCD type 5:



type 5 = type 4 + type 1 (VIEW + AD_SID_AKTUALNY)

Slowly Changing Dimension

type 6 (1/2)

- Type 6 oparty jest na type 2 (kolejne wiersze z kolejnymi wersjami) rozszerzonym o atrybuty zawierające aktualną wartość.
- Powyższy zabieg pozwala dokonywać grupowania zarówno uwzględniając historyczne wartości atrybutów (type 2) jak i aktualne wartości atrybutów.
- Zaproponowany przez inżynierów HP w 2000 roku
type 6 = type 2 + z atrybutami type 3 + modyfikowanymi zgodnie z type 1

Slowly Changing Dimension type 6 (2/2)

Rows of step: Adres zmieniony (3 rows)

^ #	KL_KLIENT_ID	KL_KOD_POCZTOWY	KL_MIASTO	KL_ULICA	klient_id	miasto	adres	kod_pocztowy
1	285	26209	Snowshoe	1336 Benin City Drive	285	Los Angeles	98 Stara Zagora Boulevard	90001
2	342	65583	Waynesville	1293 Nam Dinh Way	342	Eckert	1192 Tongliao Street	81418
3	399	77850	Concord	734 Bchar Place	399	Vera	953 Hodeida Street	74082

poprzednie wartości atrybutów

obecne wartości atrybutów

przed aktualizacją

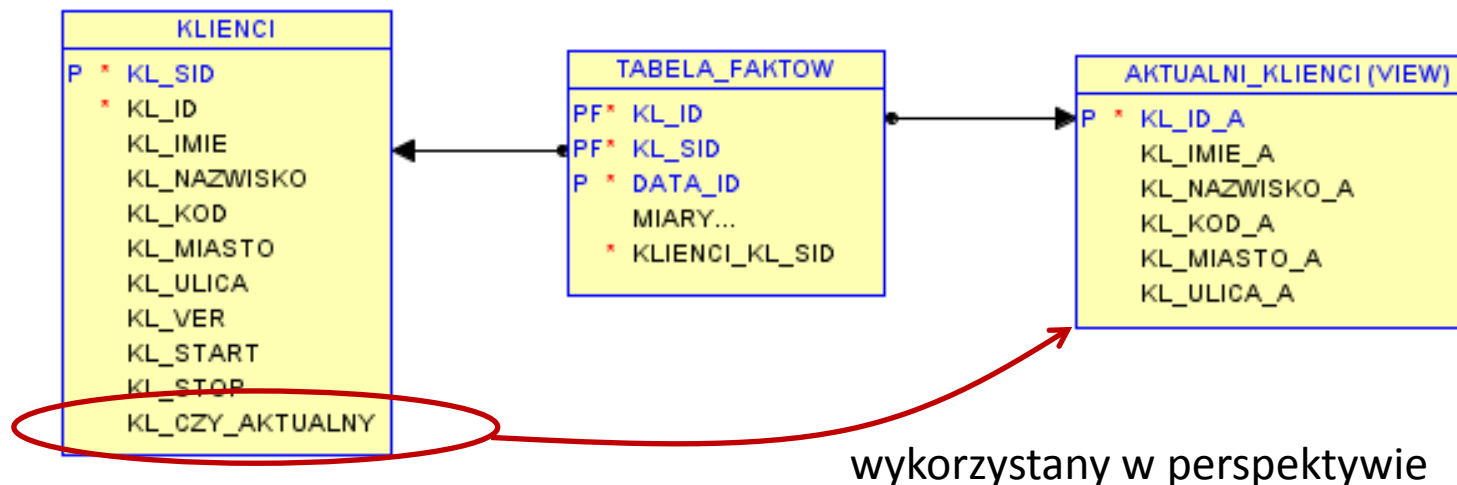
KL_ID	KL_KOD	KL_MIASTO	KL_ULICA	KL_KOD_A	KL_MIASTO_A	KL_ULICA_A	KL_SID	KL_VER	KL_START	KL_STOP
285	26209	Snowshoe	1336 Benin City Drive	26209	Snowshoe	1336 Benin City Drive	1	1	1900-01-01	2999-12-31
342	65583	Waynesville	1293 Nam Dinh Way	65583	Waynesville	1293 Nam Dinh Way	2	1	1900-01-01	2999-12-31
399	77850	Concord	734 Bchar Place	77850	Concord	734 Bchar Place	3	1	1900-01-01	2999-12-31

po aktualizacji

KL_ID	KL_KOD	KL_MIASTO	KL_ULICA	KL_KOD_A	KL_MIASTO_A	KL_ULICA_A	KL_SID	KL_VER	KL_START	KL_STOP
285	26209	Snowshoe	1336 Benin City Drive	90001	Los Angeles	28 Stara Zagora Boulevard	1	1	1900-01-01	2013-07-23
342	65583	Waynesville	1293 Nam Dinh Way	81418	Eckert	1192 Tongilao Street	2	1	1900-01-01	2013-07-23
399	77850	Concord	734 Bchar Place	74082	Vera	953 Hodeida Street	3	1	1900-01-01	2013-07-23
285	90001	Los Angeles	28 Stara Zagora Boulevard	90001	Los Angeles	28 Stara Zagora Boulevard	4	2	2013-07-23	2999-12-31
342	81418	Eckert	1192 Tongilao Street	81418	Eckert	1192 Tongilao Street	5	2	2013-07-23	2999-12-31
399	74082	Vera	953 Hodeida Street	74082	Vera	953 Hodeida Street	6	2	2013-07-23	2999-12-31

Slowly Changing Dimension type 7

- Utrzymywane są dwa wymiary
 - Type 2 – kolejne wiersze z kolejnymi wersjami
 - Type 1 – z aktualnymi wartościami atrybutów (zazwyczaj w oparciu o perspektywę na wymiarze type 2)



Slowly Changing Dimensions 5-7

wpływ na tabele faktów

- Type 5 – dodatkowy klucz do wymiaru z często zmieniającymi się atrybutami
- Type 6 – tabela faktów posiada klucz obcy oparty na sztucznym kluczu głównym tabeli wymiarów (ang. *surrogate key*)
- Type 7 – dodatkowy klucz do wymiaru z często zmieniającymi się atrybutami