

# Zadanie 3

wtorek, 29 października 2024 12:20

- Robimy teraz hurtownię danych
- PRODUKT KOŃCOWY - dokument
- 15-20 stron, wzorujemy się dokumentem z e-Nauczania
- Miara = metryka + f. agregująca
- Proces biznesowy z zadania 1, z faktów budujemy tabelę faktów
- Minimum jedna tabela faktów umawianie na wizyty lub klienci
- Tabela faktów - klucze obce + metryki np. ile zostało sprzedanych, żadnych opisów, boolean, same liczby
- Dokładanie wymiarów, zmienia nam się szczegółowość faktów
- Tabele wymiarów mają dane opisowe, mają mało wierszy, dużo kolumn, wartości składają się z przedziałów np. Wiek -> kategoria wiekowa, ilość przedziałów zależy od naszych zapytań, potem te wartości się filtruje
- Klucz surogatowy inkrementowany od 1 w tabeli wymiarów
- Klucz surogatowy musi być! Inaczej się oblewa, biznesowe mogą być, ale nie są one obowiązkowe
- Tabela faktów klucz główny połączenie wszystkich kluczy obcych
- Relacja wiele do jeden od tabeli faktów
- Zalecany schemat gwiazdy, nie zalecamy płatka śniegu - tabela wymiarów z inną tabelą wymiarów
- Księgarnia, w której pracownik pracuje, każde podpięcie inny sens biznesowy
- Auto połączenie - np. dany sprzedawca ma swojego szefa
- Tabela faktów nie może mieć null-a w kluczach obcych, ale metryki mogą być null-em
- Wartość unknown
- Związki - między tabelami wymiarów dowolne związki
- Wiele do wielu - autorstwo (Relacyjna baza danych), w HD tworzymy wówczas fakt, nie trzeba definiować żadnych metryk domyślna miara COUNT wówczas występuje
- SCD1 nie używamy w HD, SD2 flaga 0/1 (TEGO UŻYWAĆ), SCD3 data wystawiania krotki(MAŁO WYDAJNY, NIE STOSOWAĆ)
- TIME i DATA osobne tabele ZAWSZE, mogą występować wartości numeryczne, ale to tylko ma postać poglądową - 24 x 60 co do minuty tyle tworzymy takich tabel
- Tak samo data 365 dni w roku na ilość lat czyli u nas od 2018 do 2025
- Data ma dwa łączenia: koniec + początek
- Kamień milowy - przyjęcie pacjenta do szpitala (1 kamień) + operacja nowego pacjenta (2 kamień milowy) + wypis pacjenta ze szpitala (3 kamień), za każdym razem występuje data i czas
- HD nigdy nie updatujemy
- Update będzie tylko dotyczył pustych pól
- Tabela JUNK - tabela która zawiera atrybuty, które nie pasują nigdzie indziej, jak będzie tam za dużo rzeczy to tworzymy JUNK 1 JUNK 2 JUNK 3 itd.
- Opisujemy wszystkie dane w tabelach jak w pierwszym etapie
- DD - wymiar zdegenerowane, wymiar bez tabeli, dalej jako wymiar bez tabeli
- Jak się czegoś nie ma to zgoda o dodanie - trzeba napisać
- Definicja modelu wymiarowego - kostka, cała kostka wartość

BOOKSALE, wyciąganie kosteczek - filtrowanie

- Dwa tygodnie, we wtorek konsultacje, do piątku czas na poprawki, sobota Pan sprawdza
- Miara kalkulacyjna - miara wyliczana na podstawie innych miar, nie ma przypisanej funkcji agregującej WYJĄTEK!!!!
- Ostatni podpunkt per fakt, trzeba powtórzyć w każdym fakcie te powtarzające się !
- Tylko te miary, które są wyżej w tym ostatnim zadaniu! Jak nie będzie miar to będzie problem!