|  |  |  |
| --- | --- | --- |
| Krok | Cel | Opis |
| Wybór danych | Wybór danych interesujących z perspektywy biznesowej (zależne od interesariuszy) | W ramach wstępnego rozeznania wybieramy wszystkie dane, które mogą być wykorzystane w przyszłości do przeprowadzenia satysfakcjonującej interesariuszy analizy |
| Czas w kluczu | Wprowadzenie pewnego przedziału czasowego, w ramach którego będzie można analizować dane | Wprowadzamy czas jako klucz obcy odpowiedniej tabeli by móc w przyszłości analizować dane na odpowiedniej ziarnistości. Sam czas może dotyczyć np. chwili pomiaru, daty zakupu itd. |
| Dane powiązane | Zintegrowanie danych pochodzących z różnych źródeł w celu zapewnienia spójności | Zbieramy dane pochodzące z różnych źródeł (różnych baz danych) i zapisujemy je w jednym miejscu. |
| Ziarnistość | Określenie poziomu szczegółowości, w ramach którego będziemy analizować dane | Poziom szczegółowości może dotyczyć np. wspomnianego wcześniej czasu czyli w ramach jakiego okresu lub w jakim interwale chcemy przeanalizować dane. |
| Podsumowania | Przeprowadzenie agregacji danych ze względu na przyjętą ziarnistość | Dokonujemy pewnego rodzaju agregacji danych ze względu na przyjętą ziarnistość np. uśredniając dla miesiąca, maksimum z dnia itd. |
| Złączenia tabel źródłowych | Zebranie danych pochodzących z różnych tabel, ustalenie obligatoryjności wybranych złączeń | Dokonujemy złączenia danych, które są do tego momentu rozproszone na wiele tabel. Sami ustalamy jakich dokonać złączeń i co jest nam potrzebne do dalszej analizy. |
| Tworzenie tabel wynikowych | Utworzenie tabel wymiarów i faktów | Na podstawie złączonych tabel tworzymy tabele wymiarów i faktów, gdzie fakty będą konkretnymi wartościami/pomiarami następnie analizowanymi ze względu na wybrane wymiary. |
| Segregacja | Optymalizacja zapytań | Wygenerowanie obiektów pozwalających na zoptymalizowanie działania hurtowni np. indeksów, partycji, perspektyw, wyzwalaczy czy procedur ETL |

Kajetan Pynka, 254495