# Dokumentacja: Projekt Modelu Analitycznego i Zasilania

1. Informacje o projekcie

- Nazwa projektu: Hurtownia Danych BikeStores\_Analytics

- Cel projektu: Stworzenie modelu analitycznego oraz procesów zasilania danych w celu analizy sprzedaży w BikeStores.

2. Zakres dokumentacji

- Model analityczny: Projekt schematu danych hurtowni.

- Proces ETL: Szczegóły dotyczące ekstrakcji, transformacji i ładowania danych.

- Technologie: Narzędzia i technologie użyte w projekcie.

- Optymalizacje: Metody optymalizacji bazy analitycznej.

3. Analiza bazy relacyjnej pobranej z pliku `BikeStores.bak`

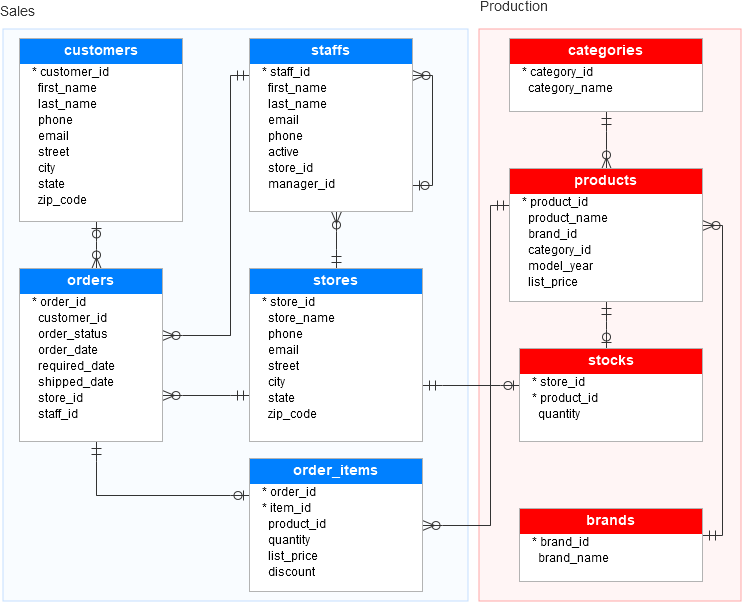
3.1. Informacje o bazie relacyjnej

- Plik ` BikeStores.bak`: Pełna kopia zapasowa bazy relacyjnej BikeStores, zaimportowana w celu zrozumienia struktury danych

- Cel analizy: Zbadanie struktury, zależności oraz danych w celu zaprojektowania hurtowni danych.

- Typ danych: Dane sprzedaży, produktów, klientów, zamówień oraz sklepów.

3.2. Struktura bazy relacyjnej



3.2.1. Tabele i ich opis

Tabela **sales.orders**

Opis: Przechowuje informacje o zamówieniach.

Klucz główny: order\_id

Kolumny kluczowe:

customer\_id: Identyfikator klienta.

store\_id: Identyfikator sklepu.

order\_date: Data zamówienia.

Tabela **sales.order\_items**

Opis: Szczegóły pozycji zamówień.

Klucz główny: order\_item\_id.

Kolumny kluczowe:

order\_id: Identyfikator zamówienia.

product\_id: Identyfikator produktu.

quantity: Ilość zamówiona.

list\_price: Cena jednostkowa.

discount: Rabat.

Tabela **production.products**

Opis: Dane o produktach.

Klucz główny: product\_id.

Kolumny kluczowe:

product\_name: Nazwa produktu.

brand\_id: Identyfikator marki.

category\_id: Identyfikator kategorii.

Tabela **production.categories**

Opis: Kategorie produktów.

Klucz główny: category\_id.

Kolumny kluczowe:

category\_name: Nazwa kategorii.

Tabela **sales.customers**

Opis: Dane klientów.

Klucz główny: customer\_id.

Kolumny kluczowe:

first\_name, last\_name: Imię i nazwisko klienta.

email: Adres e-mail.

phone: Numer telefonu.

Tabela sales.stores

Opis: Informacje o sklepach.

Klucz główny: store\_id.

Kolumny kluczowe:

store\_name: Nazwa sklepu.

city, state: Lokalizacja sklepu.

Tabela **production.brands**

Opis: Informacje o markach produktów.

Klucz główny: brand\_id.

Kolumny kluczowe:

brand\_name: Nazwa marki.

Tabela **production.stocks**

Opis: Przechowuje informacje o dostępnych zapasach produktów w sklepach.

Klucz główny: Brak (tabela łączy produkty i sklepy).

Kolumny kluczowe:

store\_id: Identyfikator sklepu.

product\_id: Identyfikator produktu.

quantity: Ilość dostępnych produktów w sklepie.

Tabela **sales.staffs**

Opis: Informacje o pracownikach.

Klucz główny: staff\_id.

Kolumny kluczowe:

first\_name, last\_name: Imię i nazwisko pracownika.

email: Adres e-mail pracownika.

phone: Numer telefonu pracownika.

store\_id: Sklep, w którym pracuje pracownik.

manager\_id: Identyfikator menedżera pracownika.

3.2.2. Relacje w bazie relacyjnej

- Klucze obce:

- `sales.orders.customer\_id → sales.customers.customer\_id`

- `sales.orders.store\_id → sales.stores.store\_id`

- `sales.staffs.store\_id → sales.stores.store\_id`

- `sales.order\_items.order\_id → sales.orders.order\_id`

- `sales.order\_items.product\_id → production.products.product\_id`

- `production.products.category\_id → production.categories.category\_id`

- `production.products.brand\_id → production.brands.brand\_id`

- `production.stocks.store\_id → sales.stores.store\_id`

- `production.stocks.product\_id → production.products.product\_id`

- `sales.staffs.manager\_id → sales.staffs.staff\_id`

4. Architektura i model analityczny



4.1. Struktura hurtowni danych

- Schemat gwiazdy:

Hurtownia oparta jest na modelu gwiazdy, składającym się z:

- Tabela faktów:

- `Fact\_Sales`: Przechowuje szczegóły transakcji, takie jak ilość, rabaty, wartość sprzedaży, oraz identyfikatory wymiarów (np. `product\_id`, `customer\_id`, `date\_id`).

- Tabele wymiarów:

- `Dim\_Products`: Dane o produktach (nazwa, kategoria, marka).

- `Dim\_Customers`: Dane klientów (imię, nazwisko, lokalizacja).

- `Dim\_Stores`: Informacje o sklepach (nazwa, lokalizacja).

- `Dim\_Staffs`: Dane o pracownikach

- `Dim\_Categories`: Dane o kategoriach produktó, identyfikatory i nazwy kategorii

- `Dim\_Time`: Dane kalendarzowe

4.2. Relacje między tabelami

Tabela faktów łączy się z tabelami wymiarów za pomocą kluczy obcych:

- `Fact\_Sales.product\_id → Dim\_Products.product\_id`

- `Fact\_Sales.customer\_id → Dim\_Customers.customer\_id`

- `Fact\_Sales.store\_id → Dim\_Stores.store\_id`

- `Fact\_Sales.date\_id → Dim\_Time.date\_id`

5. Proces ETL

5.1. Ekstrakcja danych

- Źródła danych:

- System operacyjny `BikeStores` (baza relacyjna).

- Pliki CSV z danymi historycznymi.

- Narzędzia: SQL Server Integration Services (SSIS) / Python (biblioteka `pandas`).

5.2. Transformacja danych

- Czyszczenie danych:

- Usunięcie duplikatów.

- Uzupełnianie brakujących wartości (`NULL` → wartości domyślne).

- Agregacja:

- Grupowanie sprzedaży według daty, produktu i sklepu.

- Mapowanie wymiarów:

- Łączenie danych transakcyjnych z tabelami wymiarów.

5.3. Ładowanie danych

- Ładowanie do tabel faktów i wymiarów:

- Tabele wymiarów są ładowane jako pierwsze, aby zagwarantować poprawne klucze obce.

- Tabela faktów jest ładowana z użyciem kluczy obcych.

6. Szczegóły modelu analitycznego

6.1. Schemat tabel

- Tabela: `Fact\_Sales`

- Kolumny:

- `sale\_id`: Klucz główny.

- `product\_id`, `customer\_id`, `store\_id`, `date\_id`: Klucze obce.

- `quantity`: Ilość sprzedanych sztuk.

- `total\_sales`: Wartość sprzedaży.

- `discount`: Zastosowany rabat.

- Tabela: `Dim\_Products`

- Kolumny:

- `product\_id`: Klucz główny.

- `product\_name`, `category\_id`, `brand\_id`.

- Tabela: `Dim\_Customers`

- Kolumny:

- `customer\_id`: Klucz główny.

- `first\_name`, `last\_name`, `email`, `address`.

- Tabela: `Dim\_Time`

- Kolumny:

- `date\_id`: Klucz główny.

- `date`, `year`, `month`, `week`, `weekday\_name`, `is\_holiday`.

6.2. Widoki i agregacje

Przykład widoku analitycznego:

CREATE VIEW vw\_top\_products AS

SELECT

p.product\_name,

SUM(f.quantity) AS total\_quantity,

SUM(f.total\_sales) AS total\_revenue

FROM Fact\_Sales AS f

JOIN Dim\_Products AS p ON f.product\_id = p.product\_id

GROUP BY p.product\_name

ORDER BY total\_revenue DESC;

7. Optymalizacje

- Indeksy: Na kluczach połączeń (`product\_id`, `store\_id`, `date\_id`).

- Partycjonowanie: Tabeli `Fact\_Sales` według kolumny `date\_id`.

- Widoki materializowane: Indeksowane widoki dla najczęściej wykonywanych zapytań.

- Statystyki: Regularne aktualizowanie statystyk.

8. Technologie

- Baza danych: Microsoft SQL Server.

- Narzędzia ETL: SSIS / Python.

- Analityka: Power BI, SQL Server Management Studio.

9. Harmonogram prac

- Etap 1: Projektowanie modelu analitycznego.

- Etap 2: Implementacja schematu bazy danych.

- Etap 3: Projektowanie procesów ETL.

- Etap 4: Testowanie i optymalizacja zapytań.

10. Autorzy

- Imię i nazwisko: [Twoje imię i nazwisko]

"""