Wojskowa Akademia Techniczna

Wydział Cybernetyki

Instytut Systemów Informatycznych



HURTOWNIE DANYCH

Sprawozdanie z zajęć laboratoryjnych

Wersja 01

Autorzy:

Radosław Relidzyński, Hubert Filipczuk, Marcin Dołowy, Radosław Potocki

Grupa:

WCY20IJ1S1

Podgrupa:

Grupa 4

Prowadzący laboratorium:

dr inż. Stefan Jan Rozmus

Warszawa, rok akademicki 2022/2023

|  |  |  |  |
| --- | --- | --- | --- |
| **Metryka dokumentu** | | | |
| Nazwa dokumentu | . | | |
| Opis dokumentu |  | | |
| Autorzy |  | | |
| Nazwa pliku | HD\_lab\_WCY20IJ1S1\_gr4\_v1.0\_30.01.2023.docx | Liczba stron | 21 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Historia zmian dokumentu | | | | | | |
| Wersja | Data wydania | Opis | Akcja (\*) | Rozdziały (\*\*) | Autorzy (\*\*\*) | Data kontroli |
| 01 | 2022-11-21 | Utworzenie dokumentu | N | W |  | nd |
| 1.0 | 2023-01-31 | Ostateczna wersja | We | We |  | nd |

(\*) Akcje: W = Wstaw, Z = Zamień, We = Weryfikuj, N = Nowy

(\*\*) Rozdziały: W = Wszystkie

(\*\*\*) Autorzy: patrz metryka dokumentu

**Spis treści**

[1. Opis dokumentu 5](#_Toc126075632)

[1.1. Cel dokumentu 5](#_Toc126075633)

[1.2. Wymagana wiedza 5](#_Toc126075634)

[1.3. Dokumenty powiązane 6](#_Toc126075635)

[1.4. Wykaz skrótów i oznaczeń 6](#_Toc126075636)

[2. Zawartość merytoryczna dokumentu 6](#_Toc126075637)

[2.1. Opis źródeł danych 6](#_Toc126075638)

[2.2. Wstępna lista raportów 8](#_Toc126075639)

[2.3. Model punktowy 9](#_Toc126075640)

[2.4. Model logiczny bazy danych dla hurtowni danych 11](#_Toc126075641)

[2.5. Projekt procesu ETL 11](#_Toc126075642)

[2.6. Opis kostki wielowymiarowej 12](#_Toc126075643)

[2.7. Prezentacja raportów 15](#_Toc126075644)

[2.8. Wnioski końcowe 21](#_Toc126075645)

**Spis rysunków**

[Rysunek 1 Logiczny model bazy NorthWind 7](#_Toc126075650)

[Rysunek 2. Model punktowy 9](#_Toc126075651)

[Rysunek 3. Model logiczny hurtowni. 11](#_Toc126075652)

[Rysunek 4. Projekt procesu ETL 12](#_Toc126075653)

[Rysunek 5. Struktura kostki wielowymiarowej 13](#_Toc126075654)

[Rysunek 6. Hierarchia wymiarów czasu kostki wielowymiarowej 14](#_Toc126075655)

[Rysunek 7. Widok kostki wielowymiarowej 15](#_Toc126075656)

[Rysunek 8 Raport wartości sprzedaży na wrzesień 1997 przez poszczególnych pracowników 15](#_Toc126075657)

[Rysunek 9 Raport wartości zrealizowanych zamówień klientów, którym przysługuje rabat 17](#_Toc126075658)

[Rysunek 10 Raport wartości sprzedanych jednostek danego produktu w poszczególnych państwach 18](#_Toc126075659)

[Rysunek 11 Liczba zrealizowanych zamówień przez danego dostawcę 19](#_Toc126075660)

**Spis tabel**

[Tabela 1. Lista dokumentów powiązanych 6](#_Toc126075661)

[Tabela 2. Wykaz skrótów i oznaczeń 6](#_Toc126075662)

[Tabela 3. Model punktowy – część opisowa 9](#_Toc126075663)

[Tabela 4 Zestawienie wartości sprzedanych produktów na wrzesień 1997 przez poszczególnych pracowników 16](#_Toc126075664)

[Tabela 5 Zestawienie wartości zrealizowanych zamówień klientów, którym przysługuje rabat 17](#_Toc126075665)

[Tabela 6 Zestawienie liczby sprzedanych jednostek produktu w poszczególnych państwach 18](#_Toc126075666)

[Tabela 7 Zestawienie liczby zrealizowanych zamówień przez danego dostawcę 19](#_Toc126075667)

[Tabela 8 Zestawienie najlepiej sprzedających się produktów na przestrzeni miesiąca 20](#_Toc126075668)

# Opis dokumentu

## Cel dokumentu

Celem dokumentu jest stworzenie dokumentacji technicznej hurtowni danych służącej do analizy biznesowej firmy zajmującej się wytwarzaniem produktów oraz ich dostarczaniem.

Dokument spełnia poniższe zadania:

1. Opisuje bazę danych pod względem jej struktury i sposobu organizacji danych.
2. Analizuje źródła danych pod względem jakości dostarczanych informacji.
3. Przedstawia problemy biznesowe, do których rozwiązania przydatna może się okazać analiza danych historycznych.
4. Opisuje oraz prezentuje raporty realizujące analizę funkcjonowania przedsiębiorstwa w kontekście przedstawionych problemów biznesowych.
5. Określa model punktowy prezentacji przestrzennej.
6. Przedstawia schemat relacyjnej bazy danych dla hurtowni danych.
7. Określa proces ETL przetwarzania danych do hurtowni danych.
8. Zawiera opis kostki wielowymiarowej jako źródła dla raportów.

## Wymagana wiedza

Podstawowe umiejętności wymagane do budowy hurtowni danych to:

* Zdolność do identyfikacji zależności pomiędzy danymi zwłaszcza w przypadku danych wieloźródłowych,
* Średnio zaawansowana umiejętność tworzenia oraz interpretacji zapytań SQL,
* Podstawowa wiedza z zakresu administracji bazami danych (tworzenie indeksów, backupów, vacumów itp.),

Zastosowanie hurtowni danych może być dość szerokie. W firmach handlowych może to być np. połączenie danych w raportach:

* Systemów handlowych oraz systemów analityki web Google Analytics,
* Platform ecommerce oraz systemu sprzedażowego,
* Plików od dostawców usług np. DPD i powiązanie ich z dokumentami sprzedażowymi, dzięki temu możliwe jest dokładne badanie rentowności.

## Dokumenty powiązane

Rozdział ten zawiera wykaz dokumentów, które były wykorzystywane w trakcie projektu i opracowania sprawozdania. Do wymieniania dokumentów powiązanych stosuje się tabelę, Tabela 1.

Tabela 1. Lista dokumentów powiązanych

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Nr | Nazwa | Wykaz wykorzystanych części dokumentu | Wersja | Data wydania |
|  | „Opis bazy Northwind.docx” | Opisy tabel | nd | 2022-11-09 |

## Wykaz skrótów i oznaczeń

Tabela . Wykaz skrótów i oznaczeń

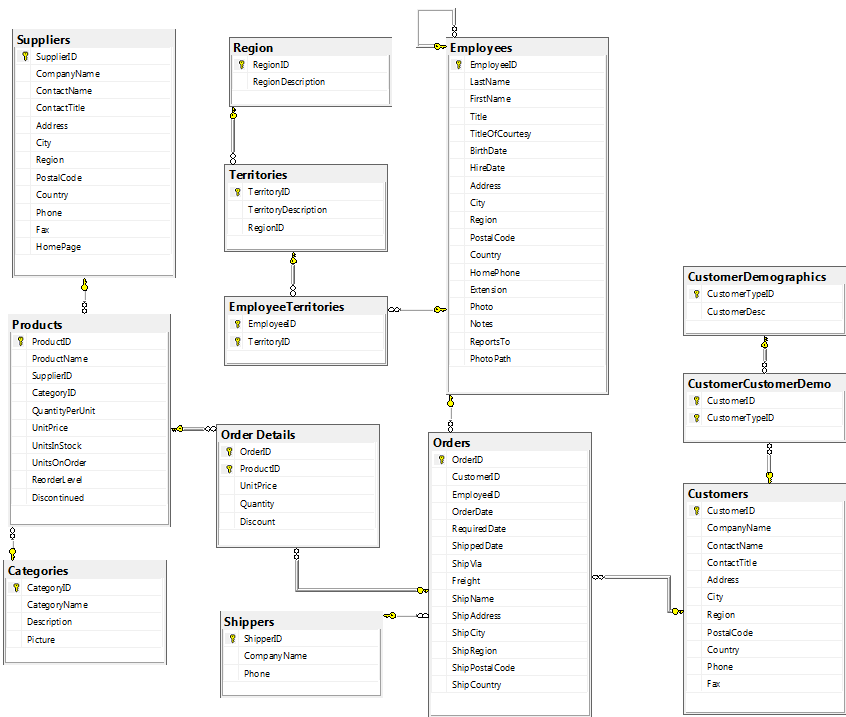
|  |  |  |
| --- | --- | --- |
| Nr | Skrót | Oznaczenie |
| S1 | SQL | Structured Query Language |
| S2 | SSMS | SQL Server Management Studio |
| S3 | ETL | Extract, Transform, Load (typ procesu) |

# Zawartość merytoryczna dokumentu

## Opis źródeł danych

Lista tabel bazy danych wraz z kompletnym opisem ich kolumn zawarty jest w dokumencie D1.

Model logiczny bazy przedstawiony został na rysunku Rysunek 1 Logiczny model bazy NorthWind.



Rysunek 1 Logiczny model bazy NorthWind

## Wstępna lista raportów

Wybór pracownika miesiąca – Który pracownik osiągnął największą sumę ze sprzedaży produktów jednostkowych w danym miesiącu? – Wartość sprzedaży produktów na przestrzeni danego miesiąca przez poszczególnego pracownika.

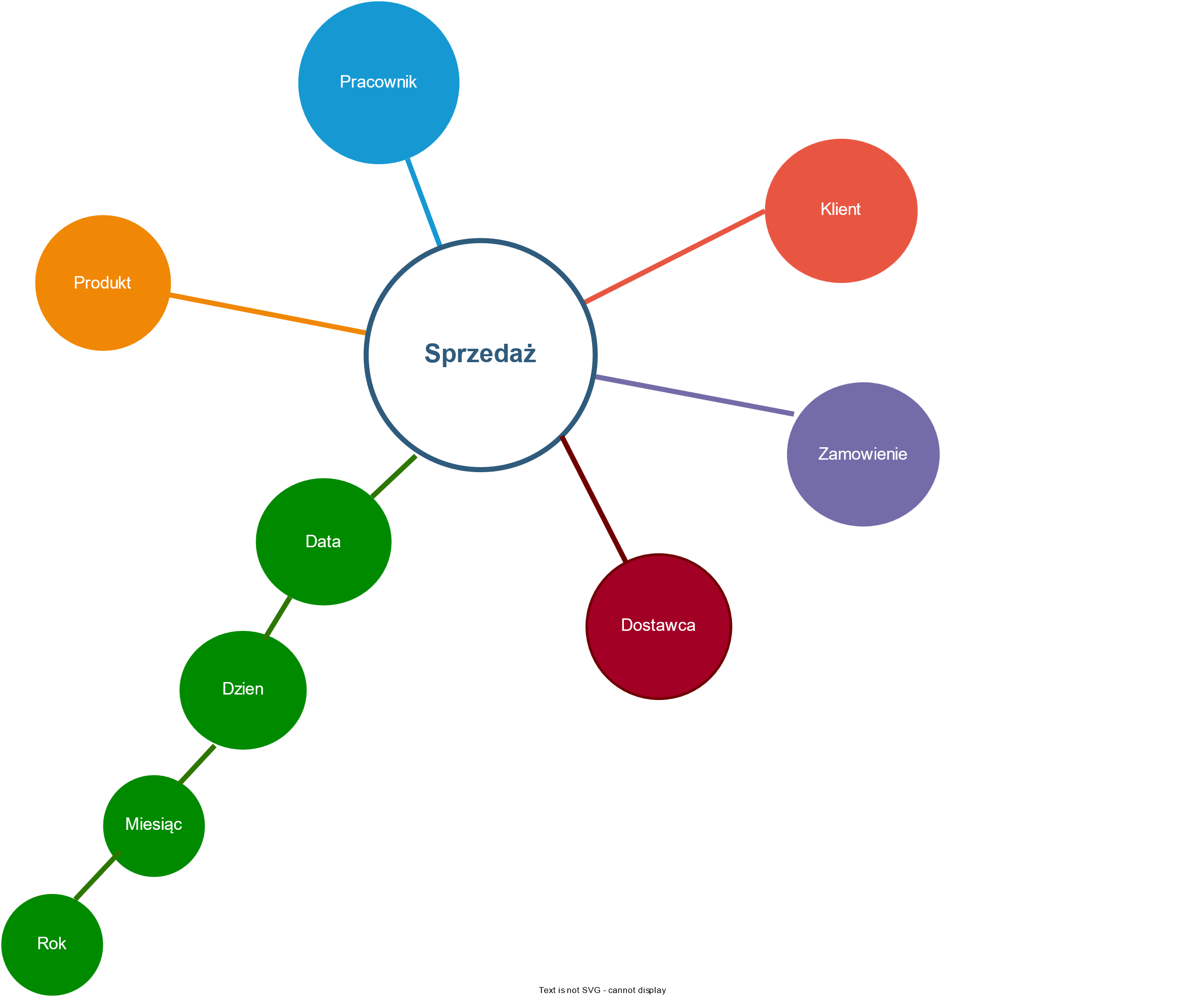
Wybór klientów, którym przysługuje zniżka – Których klientów wartość zrealizowanych zamówień przekracza minimalną kwotę, aby uzyskać rabat? – Wartość zrealizowanych zamówień przez danego klienta.

Planowanie dystrybucji produktów na poszczególne państwa – Ile jednostek danego produktu zamówił dany kraj w danym przedziale czasu? – Liczba sprzedanych jednostek produktu dla danego kraju z podziałem na przedziały czasowe.

Wybór dostawcy, z którym nie należy przedłużać umowy – Który dostawca najrzadziej brał udział przy realizacji zamówień? – Liczba zrealizowanych zamówień przez danego dostawcę.

Wybór najlepiej sprzedających się produktów - Jakie produkty wygenerowały największy obrót w danym miesiącu? – Wartość sprzedanych produktów na przestrzeni danego miesiąca.

## Model punktowy



Rysunek . Model punktowy

Tabela 3. Model punktowy – część opisowa

| **Nazwa obiektu** | **Typ** | **Zmienność w czasie** | **Uwagi** | **Typ danych** | **Wymagalność** | **Źródło** |
| --- | --- | --- | --- | --- | --- | --- |
| Sprzedaż | Encja | Tak | Zapotrzebowanie produktów przez klientów jest zmienne. Ponadto klient może anulować zamówienie | n/d | n/d | Northwind |
| Pracownik | Encja | Tak | Pracownik może okresowo rezygnować z pracy, a następnie ją wznawiać | n/d | n/d | Northwind |
| Klient | Encja | Nie | Klient może zmienić imię i nazwisko, ale nie ma to znaczenia | n/d | n/d | Northwind |
| Zamówienie | Encja | Tak | Wymagana data realizacji może ulec zmianie | n/d | n/d | Northwind |
| Dostawca | Encja | Stała | Dostawca zamówienia nie ulega zmianie | n/d | n/d | Northwind |
| Produkt | Encja | Tak | Cena jednostkowa produktu może ulec zmianie | n/d | n/d | Northwind |
| Data | Encja | Stała | Oznaczenie dnia, miesiąca i roku | n/d | n/d | Northwind |
| Dzień | Encja | Stała | Dzień określony datą kalendarzową jest wielkością stała | n/d | n/d | Northwind |
| Miesiąc | Encja | Stała | Każdy rok ma 12 miesięcy | n/d | n/d | Northwind |
| Kwartał | Encja | Stała | Każdy rok ma 4 kwartały | n/d | n/d | Northwind |
| Rok | Encja | Stała | Rok wykonania czynności nie zmienia się | n/d | n/d | Northwind |
| Dzień → Miesiąc | Relacja | Stała | Dzień określony datą kalendarzową jest wielkością stała | n/d | n/d | n/d |
| Miesiąc → Kwartał | Relacja | Stała | Relacja pomiędzy miesiącem a kwartałem jest stała | n/d | n/d | n/d |
| Kwartał → Rok | Relacja | Stała | Relacja pomiędzy miesiącem a kwartałem jest stała | n/d | n/d | n/d |
| Pracownik → Sprzedaż | Relacja | Tak | Liczba sprzedanych produktów przez pracownika jest zmienna | n/d | n/d | n/d |
| Zamówienie → Sprzedaż | Relacja | Tak | Można anulować zamówienie | n/d | n/d | n/d |
| Dostawca → Sprzedaż | Relacja | Stała | Nie możemy zmienić dostawcy, jeśli zamówienie jest w trakcie realizacji | n/d | n/d | n/d |
| Produkt → Sprzedaż | Relacja | Tak | Cena produktu może ulec zmianie | n/d | n/d | n/d |
| Sprzedaz.wartosc | Atrybut | Tak | Dopuszczalna jest modyfikacja zamówienia | money | Tak | n/d |
| Sprzedaz.ilosc | Atrybut | Tak | Dopuszczalna jest modyfikacja zamówienia | int | Tak | n/d |
| Klient.klientID | Atrybut | Stała | Identyfikator klienta nie ulega zmianie | nchar(5) | Tak | Northwind.Customers.CustomerID |
| Klient.nazwisko | Atrybut | Stała | Nazwisko klienta nie ulega zmianie | varchar(256) | Tak | Northwind.Customers.ContactName |
| Pracownik.pracownikID | Atrybut | Stała | Identyfikator pracownika nie ulega zmianie | int | Tak | Northwind.Employees.EmployeeID |
| Pracownik.nazwisko | Atrybut | Stała | Nazwisko pracownika nie ulega zmianie | varchar(256) | Tak | Northwind.Employees.LastName |
| Produkt.produktID | Atrybut | Stała | Identyfikator produktu nie ulega zmianie | int | Tak | Northwind.Products.ProductID |
| Produkt.cenaJednostkowa | Atrybut | Tak | Cena produktu może ulec zmianie | money | Tak | Northwind.Products.UnitPrice |
| Zamowienie.zamowienieID | Atrybut | Stała | Identyfikator zamówienia nie ulega zmianie | int | Tak | Northwind.Orders.OrderID |
| Zamowienie.krajZamowienia | Atrybut | Stała | Kraj zamówienia nie ulega zmianie | navchar(15) | Tak | Northwind.Orders.Shipcountry |
| Dostawca.dostawcaID | Atrybut | Stała | Identyfikator dostawcy nie ulega zmianie | int | Tak | Northwind.Shippers.CompanyName |
| Dostawca.nazwaDostawcy | Atrybut | Stała | Nazwa dostawcy nie ulega zmianie | nvarchar(40) | Tak | Northwind.Shippers.CompanyName |

## Model logiczny bazy danych dla hurtowni danych

Poniższy rysunek przedstawia model logiczny dla hurtowni danych.

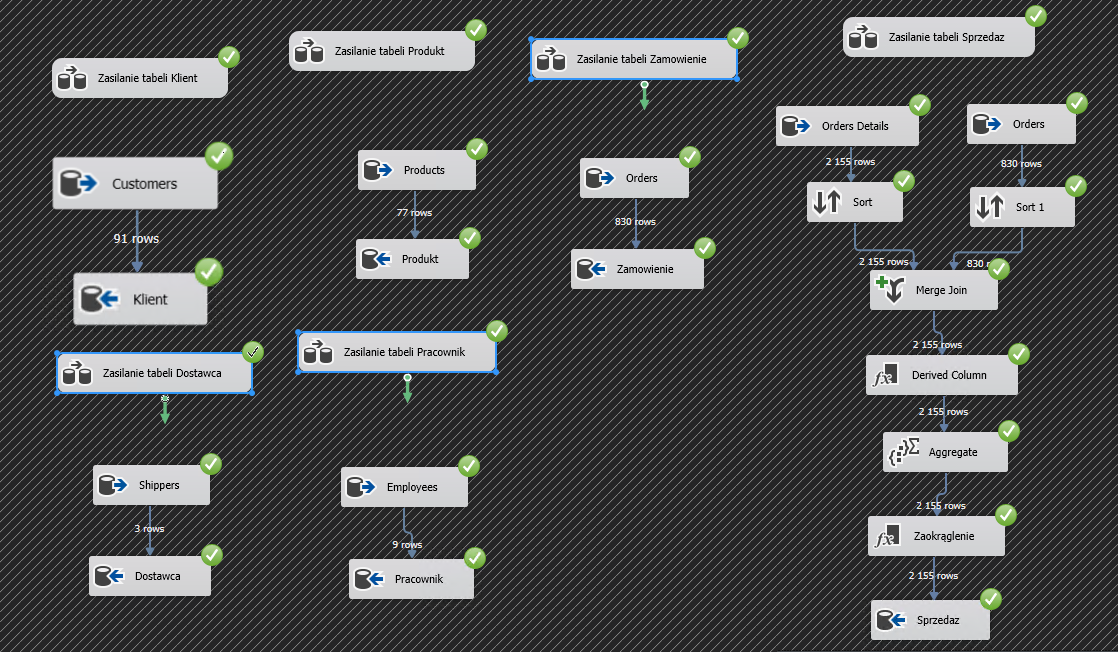
Obraz zawierający tekst, wewnątrz, zrzut ekranu

Opis wygenerowany automatycznie

Rysunek . Model logiczny hurtowni.

## Projekt procesu ETL

Proces ETL (Extract, Transform, Load) jest techniką integracji danych pochodzących z różnych źródeł, która polega na wyodrębnieniu danych (Extract), przekształceniu ich (Transform) i załadowaniu do docelowego systemu (Load). Jest to kluczowa czynność w procesie zarządzania danymi i jego cel to zapewnienie jednolitego i aktualnego dostępu do danych w organizacji. Projekt procesu ETL dla naszej hurtowni danych przedstawiony jest poniżej.



Rysunek . Projekt procesu ETL

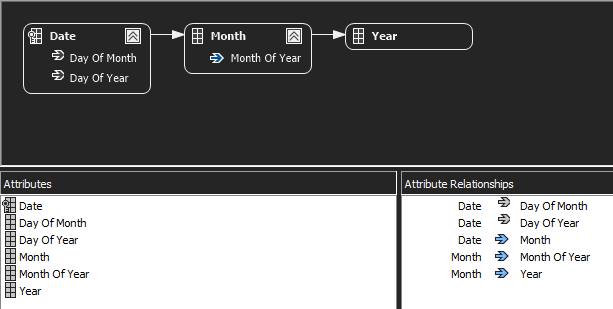
## Opis kostki wielowymiarowej

Kostka wielowymiarowa to specjalne narzędzie do przechowywania i analizy danych, które pozwala na łatwe wizualizowanie i agregację informacji. Składa się z wymiarów i faktów, gdzie wymiary opisują atrybuty danych, a fakty to wartości przypisane do poszczególnych kombinacji wymiarów. Kostki wielowymiarowe pozwalają na szybką i wygodną analizę danych, co czyni je niezbędnym narzędziem w wielu dziedzinach, takich jak biznes, nauka i analiza danych. Model kostki, widok jej przeglądania oraz hierarchia czasu zostały przedstawione na poniższych rysunkach.

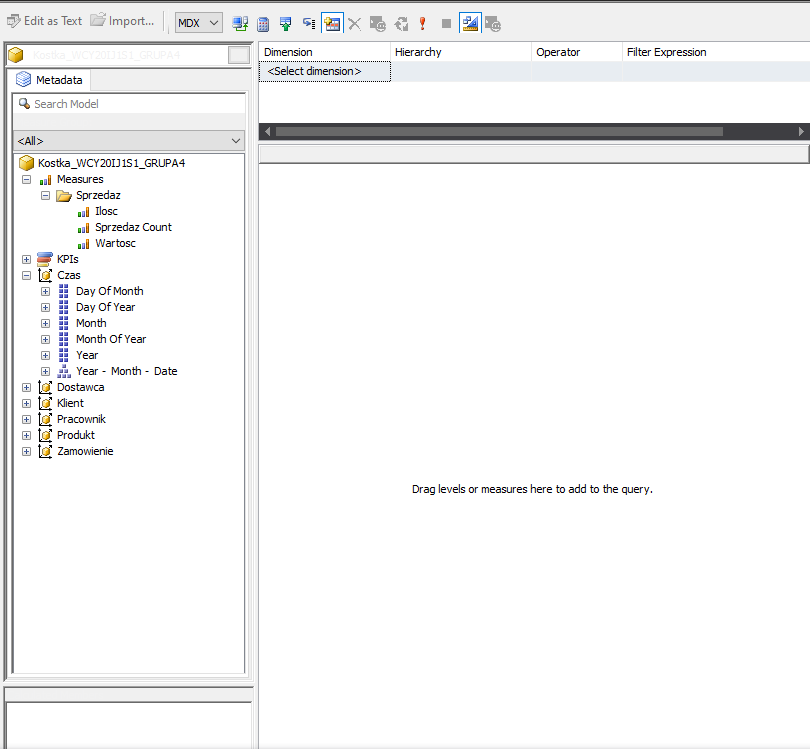
Obraz zawierający tekst, zrzut ekranu, parking

Opis wygenerowany automatycznie

Rysunek . Struktura kostki wielowymiarowej



Rysunek . Hierarchia wymiarów czasu kostki wielowymiarowej



Rysunek . Widok kostki wielowymiarowej

## Prezentacja raportów

Obraz zawierający stół

Opis wygenerowany automatycznie

Rysunek Raport wartości sprzedaży na wrzesień 1997 przez poszczególnych pracowników

Tabela Zestawienie wartości sprzedanych produktów na wrzesień 1997 przez poszczególnych pracowników

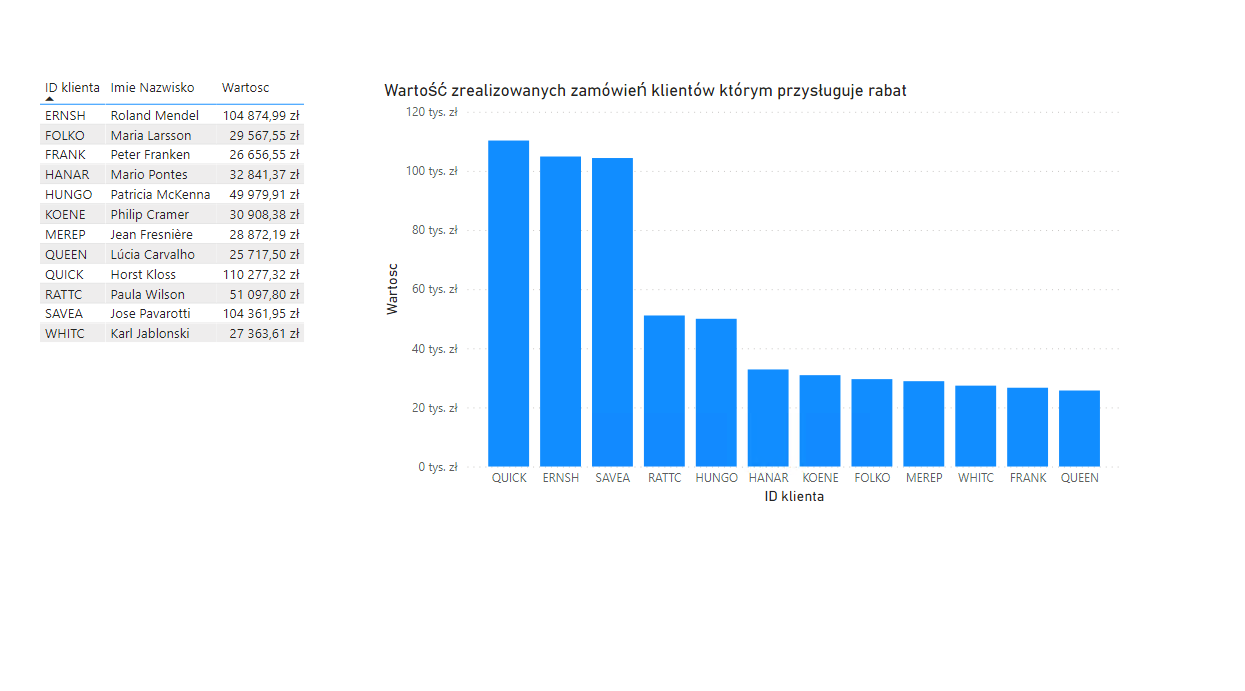
Obraz zawierający stół

Opis wygenerowany automatycznie

Powyższy raport pozwala na wybranie najlepszego pracownika danego miesiąca. Wyznaczeni zostają Ci pracownicy, którzy osiągnęli najwyższą sumę ze sprzedaży produktów jednostkowych.

Przykładowo, we wrześniu 1997 roku, największą sumę ze sprzedaży produktów jednostkowych osiągnął pracownik o id równym 4 z sumą wynoszącą 13475,99 zł.

W celu przedstawienia danych i uzyskania czytelnego i przejrzystego wyniku, w raporcie użyto tabeli oraz wykresu. Tabela pomaga w prezentacji szczegółowych danych liczbowych, natomiast wykres umożliwia łatwe i szybkie zrozumienie tendencji i zależności między poszczególnymi wartościami. Obie te formy prezentacji pozwalają na szybką analizę i wnioskowanie na podstawie prezentowanych danych.



Rysunek Raport wartości zrealizowanych zamówień klientów, którym przysługuje rabat

Tabela Zestawienie wartości zrealizowanych zamówień klientów, którym przysługuje rabat

Obraz zawierający stół

Opis wygenerowany automatycznie

Dzięki powyższemu raportowi jesteśmy w stanie zestawić klientów, którym przysługuje rabat. Rabat ten jest przyznawany klientom, których wartość zrealizowanych zamówień przekracza minimalną kwotę, w naszym przypadku wynosi ona 25000 złotych.

Rabat dla klientów, którzy wydadzą powyżej 25000 złotych, może zostać wprowadzony w celu zachęcenia do większych zakupów. Taki rabat jest doskonałym narzędziem do zwiększenia lojalności klientów i zachęcenia ich do dokonywania większych zakupów. Wzrost wydatków klientów przekłada się na wzrost przychodów dla sklepu, co jest korzystne zarówno dla klientów, jak i dla samego sklepu. Warto zauważyć, że taki rabat może również zwiększyć konkurencyjność sklepu na rynku, co jest korzystne dla jego rozwoju i stabilności finansowej.

Obraz zawierający mapa

Opis wygenerowany automatycznie

Rysunek Raport wartości sprzedanych jednostek danego produktu w poszczególnych państwach

Tabela Zestawienie liczby sprzedanych jednostek produktu w poszczególnych państwach

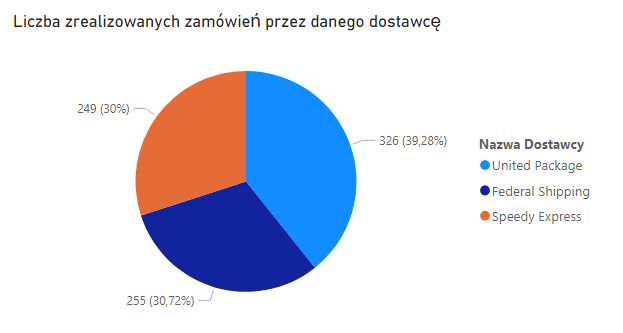
Obraz zawierający stół

Opis wygenerowany automatycznie

Powyższy raport w formie zrzutu ekranu jest mało czytelny oraz przydatny, natomiast w programie jako interaktywna mapa jest efektywny, przejrzysty oraz intuicyjny. Mapa jest bardzo użytecznym narzędziem w prezentacji danych geograficznych, takich jak sprzedaż produktów. Użycie mapy w raporcie pozwala na wizualną prezentację danych, co umożliwia szybką i łatwą analizę. Mapa pozwala również na łatwe porównanie wartości sprzedaży produktu w poszczególnych regionach i wyciągnięcie wniosków dotyczących trendów i zależności między sprzedażą a lokalizacją.

Dzięki mapie można również identyfikować mocne i słabe strony poszczególnych regionów, co pozwala na lepsze zrozumienie rynku i podejmowanie lepiej ukierunkowanych działań.

W naszym przypadku raport ten przedstawia liczbę sprzedanych jednostek produktu o id równym 3 w poszczególnych państwach. Widzimy, że głównie produkt trafiał do Wenezueli, Niemiec oraz Szwecji. Mniejsze zamówienia trafiały do Stanów Zjednoczonych oraz Danii.



Rysunek Raport liczby zrealizowanych zamówień przez danego dostawcę

Tabela Zestawienie liczby zrealizowanych zamówień przez danego dostawcę

Obraz zawierający stół

Opis wygenerowany automatycznie

Raport porównujący dostawców na podstawie liczby zrealizowanych zamówień w ciągu trzech lat jest bardzo przydatnym narzędziem, które pozwala na dokładną analizę i ocenę wydajności dostawców. Zawartość raportu jest niezwykle przydatna dla osób odpowiedzialnych za zarządzanie biznesem, ponieważ pozwala na łatwe i szybkie porównanie wydajności dostawców w określonym czasie.

Dzięki raportowi można łatwo identyfikować najlepiej i najgorzej funkcjonujących dostawców, co pozwala na wyciągnięcie wniosków dotyczących ich działalności i na lepsze zrozumienie rynku. Raport pozwala także na zidentyfikowanie potencjalnych trendów i zmian w działalności dostawców, co jest niezwykle istotne w procesie podejmowania decyzji biznesowych.

Warto również zauważyć, że taki raport jest bardzo przydatny dla osób odpowiedzialnych za negocjacje z dostawcami. Dzięki dokładnej analizie wydajności dostawców, można lepiej uzgodnić warunki współpracy i negocjować lepsze ceny, co jest kluczowe dla rentowności biznesu.

W skrócie, raport porównujący dostawców na podstawie liczby zrealizowanych zamówień jest bardzo przydatnym narzędziem dla wszystkich osób odpowiedzialnych za zarządzanie biznesem. Pozwala na dokładną analizę wydajności dostawców i na podjęcie ukierunkowanych działań, co jest kluczowe dla sukcesu biznesu.

Tabela Zestawienie najlepiej sprzedających się produktów na przestrzeni miesiąca

Obraz zawierający stół

Opis wygenerowany automatycznie

Raport ten umożliwia porównanie wydajności poszczególnych produktów na przestrzeni jednego miesiąca, co pozwala na ocenę które produkty były najbardziej popularne i wygenerowały największy obrót. Dzięki temu raportowi można identyfikować najlepiej sprzedające się produkty i skupić się na ich dalszej promocji lub zastanowić się nad dodaniem do oferty nowych, potencjalnie popularnych produktów.

Warto również zwrócić uwagę na produkty, które wygenerowały mniejsze obroty i zastanowić się nad przyczyną tego stanu rzeczy, np. brakiem dostępności, niską jakością lub brakiem zainteresowania ze strony klientów.

Podsumowując, raport ten pomaga w uzyskaniu pełnego obrazu sprzedaży poszczególnych produktów i umożliwia wyciągnięcie wniosków dotyczących przyszłej strategii biznesowej.

## Wnioski końcowe

Hurtownia danych to centralne miejsce, w którym zgromadzone są dane z różnych źródeł, takich jak systemy transakcyjne, dane demograficzne, dane o klientach i dane o produktach. Celem jest umożliwienie szybkiego i łatwego dostępu do danych i umożliwienie tworzenia raportów i analiz danych.

Przekształcenie bazy danych w hurtownię danych jest korzystne z wielu powodów. Po pierwsze, pozwala na łatwy i szybki dostęp do danych, co umożliwia szybsze podejmowanie decyzji biznesowych. Po drugie, umożliwia tworzenie raportów na temat danych firmy, co pozwala na lepsze zrozumienie jej sytuacji finansowej i funkcjonowania.

Proces tworzenia hurtowni danych był cenną lekcją, która pogłębiła naszą wiedzę na temat procesów ETL, kostek wielowymiarowych, tworzenia raportów oraz utrwaliła nasze umiejętności dotyczące baz danych. To doświadczenie pozwala nam dostrzegać wiele korzyści, jakie płyną z posiadania dobrze zaprojektowanej hurtowni danych, a także skutecznie wykorzystywać ją w codziennej pracy.

Podsumowując, hurtownia danych to narzędzie, które pozwala na ulepszenie procesu zarządzania danymi i podejmowania decyzji biznesowych. Dzięki niej, można uzyskać szybki i łatwy dostęp do danych, a także tworzyć raporty i analizy, co pozwala na lepsze zrozumienie sytuacji firmy i funkcjonowania rynku.