

Michał Kukliński

Budowa i funkcjonowanie hurtowni danych w kontekście efektywności podejmowania decyzji gospodarczych

Słowa kluczowe: *hurtownia danych, business intelligence, data mining, data warehouse*

Abstrakt: W skomputeryzowanych przedsiębiorstwach, gromadzących ogromne ilości danych, ich przetwarzanie w tradycyjny sposób będzie wysoce nieefektywne i nie dostarczy tak wiele interesujących informacji, prognoz i relacji, jak systemy Business Intelligence, których hurtownie danych są podstawą. Publikacja jest odpowiedzią na pytania: co to jest hurtownia danych, do czego służy. Przedstawione zostały etapy budowy hurtowni danych oraz czynniki zapewniających osiągnięcie sukcesu w podejmowaniu decyzji gospodarczych. Przykłady zastosowania dobrze zaprojektowanej hurtowni danych ilustrują możliwość jednoczesnego funkcjonowania w kilku obszarach działalności gospodarczej. Nakłady na wprowadzenie nowej technologii wykorzystywania wiedzy zawartej w danych ponoszone są tylko raz, natomiast zwrot z inwestycji następuje równocześnie w wielu etapach efektywnego podejmowania decyzji w przedsiębiorstwie.

Wprowadzenie

Powszechnie akceptowana definicja hurtowni danych, rozwinięta przez Inmona w 1980 roku, brzmi: „skoncentrowane na temat, zintegrowane, zmienne w czasie, nieulotne gromadzenie danych wykorzystywane do podejmowania decyzji strategicznych”¹. Informacja jest jednym z głównych czynników uzyskiwania przewagi konkurencyjnej przedsiębiorstwa, dlatego każdy podmiot gospodarczy maksymalizujący zysk powinien dążyć do efektywnego wykorzystywania wiedzy zgromadzonej w danych. Hurtownia danych

¹ W. H. Inmon, *Building the Data Warehouse, Fourth Edition*, Wiley Publishing, Inc., Indianapolis 2005, s. 29.

dostarcza historycznych informacji, które są stałe, zgodne i niezawodne. Dane zgromadzone w hurtowni mogą osiągać olbrzymie rozmiary (nawet 100 terabajtów i więcej), dlatego od początku musimy to przewidzieć, umożliwiając wzrost danych w wydajny sposób oraz używając wspólnych reguł biznesowych tak, aby zgromadzone dane mogły być użyteczne dla całego przedsiębiorstwa. Dlatego w dobie skomputeryzowanych przedsiębiorstw pozyskujących ogromne ilości danych przetwarzanie ich w tradycyjny sposób byłoby wysoce nieefektywne i nie dostarczy nam tyle interesujących informacji, prognoz oraz zależności, co systemy Business Intelligence, których fundamentem są hurtownie danych. Poniższy materiał odpowie nam na pytania: czym jest hurtownia danych?, do czego służy i jakie są przykłady jej zastosowań. Poznamy etapy budowy hurtowni danych. W dalszej części publikacji przedstawię czynniki zapewniające osiągnięcie sukcesu.

Co to jest hurtownia danych?

Hurtownia danych to analityczna baza danych wykorzystywana jako podstawa systemu wspomagającego podejmowanie decyzji. Hurtownie projektuje się dla dużej liczby stałych danych².

Hurtownia danych (*data warehouse*, DW) jest zestawem narzędzi pozwalających kierownikom, dyrektorom i analitykom szybciej i skuteczniej podejmować decyzje. Podstawowym celem jest dostarczenie właściwych informacji, właściwym kosztem, we właściwym miejscu i czasie, by na ich podstawie można było podejmować właściwe decyzje. Hurtownie danych stanowią platformę integrującą dane pochodzące z różnych źródeł i umożliwiającą ich przetwarzanie analityczne (*online analytic processing*, OLAP). Są elementem pośrednim między przetwarzaniem transakcyjnym a przetwarzaniem analitycznym³.

Hurtownię danych definiujemy jako bazę danych charakteryzującą się czterema poniższymi cechami.

Zorientowanie na temat oznacza, że zbierane dane dotyczą pewnego tematu (np. sprzedaży), a nie działań (np. obsługiwanie zamówień). Operacyjne bazy danych skupiają się na procesach biznesowych; są zorientowane na działanie⁴.

² V. Poe, P. Klauer, S. Brobst, *Tworzenie hurtowni danych*, Wydawnictwa Naukowo-Techniczne, Warszawa 2000, s. 44.

³ M. Jarke, M. Lenzerini, Y. Vassiliou, P. Vassiliadis, *Hurtownie danych. Podstawy organizacji i funkcjonowania*, WSiP, Warszawa 2003, s. 15.

⁴ C. Imhoff, N. Gallemmo, J. G. Geiger, *Mastering Data Warehouse Design – Relational and Dimensional Techniques*, Wiley Publishing, Inc., Indianapolis 2003. s. 13.

Nieulotność oznacza, że dane, raz umieszczone w hurtowni, zazwyczaj pozostają niezmienione. Każdy użytkownik bazy danych ma pewność, że zapytanie zawsze zwróci taki sam wynik, niezależnie od tego, jak często jest wykonywane.

Zintegrowanie oznacza, że dane są jednolite. Na przykład daty przechowywane są zawsze w tym samym formacie. Jednolitość jest problemem dla wielu przedsiębiorstw, zwłaszcza tych, które używają różnych typów narzędzi. Niektóre różnice dotyczą kwestii podstawowych, jak choćby kodowania znaków. Zanim dane zostaną umieszczone w hurtowni danych, muszą zostać ujednolicone.

Zmienność w czasie oznacza, że gromadzone są dane zmieniające się w czasie. Prawie wszystkie zapytania kierowane do hurtowni danych wymagają prześledzenia jakiegoś odcinka czasu⁵.

Po co hurtownia danych?

Wiedza o tym, dlaczego hurtownia danych jest potrzebna, ułatwia jej projektowanie i pozwala przekonać do niej pracowników firmy. Główne korzyści ze stworzenia hurtowni danych są następujące.

1. Zintegrowany i całościowy obraz przedsiębiorstwa

Hurtownia danych łączy informacje z zasadniczo odmiennych systemów operacyjnych w jeden pełny obraz przedsiębiorstwa. Wspólne struktury, zintegrowana organizacja danych i ujednolicony zbiór definicji danych dostarczają środków umożliwiających wspomaganie podejmowania decyzji nie do osiągnięcia w rozproszonych systemach transakcyjnych. Otrzymany w ten sposób pełny obraz przedsiębiorstwa daje analitykom informacje niezbędne do podjęcia trafnych decyzji dotyczących zagadnień strategicznych.

2. Dostęp do archiwalnej informacji o przedsiębiorstwie

Zrozumienie dotychczasowych trendów i zachowań, a następnie poznanie, w jaki sposób aktualne otoczenie biznesowe wynikało z przeszłości, umożliwiają firmom opracowanie planów analiz ilościowych w celu podejmowania decyzji.

⁵ C. Todman, *Projektowanie hurtowni danych. Zarządzanie kontaktami z klientami (CRM)*, Wydawnictwa Naukowo-Techniczne, Warszawa 2003, s. 35–37.

3. Wewnątrz przedsiębiorstwa jednoznaczne źródło informacji, które jest godne zaufania

Hurtownia danych wykorzystująca zintegrowane dane i definicje zapewnia jednoznaczne, godne zaufania źródło informacji, które umożliwia podejmującym decyzje jednolite rozumienie prowadzonego w przedsiębiorstwie biznesu. Zbyt często, gdy przedsiębiorstwo wykorzystuje niezintegrowane systemy raportowania, różne osoby podejmujące decyzje mają różne wersje „liczb” charakteryzujących działalność tego przedsiębiorstwa.

4. Wspomaganie podejmowania decyzji bez ingerencji w systemy operacyjne

Systemy wspomaganie podejmowania decyzji są znane z tego, że pochłaniają ogromne ilości środków. Realizacja skomplikowanego zapytania, które najczęściej polega na przeglądaniu, łączeniu i gromadzeniu wiele milionów wierszy relacyjnych danych, stwarza ogromne zapotrzebowania na pamięć, czas pracy procesora i urządzenia wejścia/wyjścia. Rozwiązaniem jest oddzielenie na poziomie sprzętowym bazy danych dla hurtowni od operacyjnej bazy danych, tak aby użytkownicy mogli podejmować decyzje, kiedy tylko chcą, bez wpływu na bieżące operacje biznesowe.

Zastosowania Hurtowni danych

Najbardziej oczywiste korzyści dają zastosowania, które maksymalizują dochodowość firmy w sposób umożliwiający jej zmierzenie i odniesienie bezpośrednio do wyników finansowych firm. Hurtownie danych, które to umożliwiają, dostarczają informacji mających istotny wpływ na działalność przedsiębiorstwa. Poniższe przykłady pokazują, gdzie i jak różne zastosowania hurtowni danych zostały korzystnie wykorzystane w różnych gałęziach przemysłu.

1. Wykrywanie oszustw

Pomysł polega na opracowaniu, na podstawie możliwości hurtowni danych, algorytmu wykrywania oszustw, a następnie wykorzystania go w systemach operacyjnych. Zapytania *ad hoc* i możliwości hurtowni danych wspomaganie podejmowania decyzji pozwalają na zbadanie danych w celu zidentyfikowania najważniejszych obszarów, na których trzeba skoncentrować wysiłki, żeby wykryć oszustwa. Wykorzystując dane archiwalne z hurtowni

tworzy się modele do przewidywania schematów oszustw, tak aby można je było wykryć przed wystąpieniem strat.

2. Ukierunkowany marketing

Dotychczas firmy poprawiały skuteczność marketingu przez pozyskiwanie większej liczby nazwisk i kontaktując się z większą liczbą klientów. Obecnie jakość, a nie liczba klientów, z którymi należy się kontaktować, stanowi istotną różnicę. Wiedzy na temat demografii klientów, zachowań i zapotrzebowań na produkt można użyć do ukierunkowanego marketingu. Korzystając z archiwalnych informacji z hurtowni danych do opracowania modeli odpowiedzi, analitycy mogą przewidzieć prawdopodobieństwa reakcji na marketing telefoniczny lub wysyłkowy, przed rozpoczęciem marketingu. Analitycy mogą wykorzystać te prawdopodobieństwa i oczekiwaną wartość transakcji (która może być także wymodelowana za pomocą hurtowni) w celu wybrania właściwych klientów do bezpośredniego marketingu.

3. Analiza rentowności

Możliwe staje się poznanie zależności rentowności od konkretnych grup klientów oraz powiązanie wyników rentowności bezpośrednio z zachowaniami klientów. Zdolność wspomagania decyzji gwarantowana przez hurtownię danych umożliwia wgląd w to, w jaki sposób najlepiej oddziaływać na klientów, żeby zapewnić największe zyski.

4. Zatrzymywanie klientów

Program zatrzymywania klienta może polegać na specjalnych promocjach, telefonach lub listach z wyrazami uznania, lub też na akcji marketingowej mającej na celu uświadomienie i przypomnienie zalet oferty firmy. Oczywiście takie programy kosztują, więc muszą być dobrze ukierunkowane.

Dokładność w określaniu grup ryzyka zależy od jakości modeli oceny ryzyka. Jakość tych modeli opiera się na jakości i dostępności szczegółowej informacji o klientach. Informacje o klientach muszą zawierać dużo danych o ich zachowaniu, aby mogły być podstawą modeli przewidujących przyszłość.

5. Zarządzanie zapasami

Dla produktów, na które popyt jest zmienny, sprzedawcy stosują technikę zwaną sprzedażą zgodną z modą. Polega to między innymi na przenoszeniu produktów ze sklepów, gdzie popyt jest mały, do sklepów, gdzie jest on

większy. Już dawno zauważono, że moda zaczyna się w dużych miastach, a następnie rozpowszechnia się w mniej zurbanizowanych regionach. Różnice w popycie zależą od demografii klientów, różnic klimatycznych (szczególnie dotyczy to odzieży) itd. Wykorzystując informacje z hurtowni danych i technikę sprzedaży zgodną z modą, handlowcy mogą zmaksymalizować zyski na danym produkcie, póki jest na niego duży popyt, przemieszczając zapasy z jednych sklepów do innych, w zależności od miejsca zwiększonego popytu.

6. Analiza ryzyka kredytowego

W celu dokładnego określenia ryzyka związanego z udzieleniem kredytu każda większa firma kredytowa opracowała modele ilościowe, oparte na danych z przeszłości, umożliwiające przewidywanie zachowania się osobników podczas spłacania kredytu.

7. Długoterminowa ocena wartości klienta

Modele predykcyjne, używane do takiej oceny, są mocno oparte na dostępie do informacji z hurtowni danych o zachowaniu się klienta w odniesieniu do jego cech psychologicznych, demograficznych oraz do historii życia. Zachowanie klienta śledzi się w hurtowni danych poprzez takie wydarzenia, jak zakupy, ankiety i innego rodzaju kontakty z firmą. Na obraz psychologiczny klienta składają się skłonności do robienia zakupów, profile ryzyka, jego własny wizerunek itd. Dane te zdobywa się ze źródeł zewnętrznych. Do ważnych informacji z życia klienta należą: ślub, narodziny dzieci, zakup domu, pójście dzieci na studia, emerytura itp. Na koniec, informacje demograficzne to informacje o pochodzeniu etnicznym klienta, jego poziomie dochodów, zawodzie i wiele innych, które są specyficzne dla danej osoby i przewidywalne w długoterminowym okresie.

8. Ustalanie ceny

Ustalanie konkurencyjnej ceny zawiera w sobie rozpoznanie popytu na produkt, konkurencyjności rynku i granic rentowności. Wszystko to powinno być możliwe do uzyskania z hurtowni danych. Popyt na produkt można określić na podstawie danych z przeszłości, z uwzględnieniem warunków rynkowych i sezonowości. Jak napisano wcześniej, określenie elastyczności cenowej jest podstawą do podjęcia właściwych decyzji co do ceny produktu i wielkości podaży. Ustalanie konkurencyjnej ceny polega także na zbadaniu cen produktów zastępczych i alternatywnych wyborów dokonywanych przez klientów.

Etapy budowy i rozwoju hurtowni danych

Hurtownie danych są konstruowane w celu zaspokajania potrzeb informacyjnych decydentów, dlatego zawartość hurtowni powinna odzwierciedlać ich zapotrzebowanie na informacje analityczne.

Metoda projektowania hurtowni danych powinna zapewniać systematyczne uwzględnianie zmian potrzeb informacyjnych w strukturze hurtowni, a także umożliwiać rozszerzanie jej tematyki.

Budowa hurtowni danych obejmującej podstawowe obszary działania firmy jest procesem czasochłonnym i może trwać nawet kilka lat. Wybrana metoda rozwoju hurtowni musi więc łączyć w sobie możliwość rozwijania wizji całości, ale także szybkie uzyskiwanie efektów na poszczególnych etapach.

Projektowanie hurtowni powinno być oparte na następujących zasadach:

1) stopniowy rozwój hurtowni danych tak, aby efekty cząstkowe były widoczne w krótkim okresie. Fragmenty, odpowiedzialne za wydzielone problemy decyzyjne już funkcjonują i są wykorzystywane, a jednocześnie stanowią stymulację dla kolejnych użytkowników w zakresie zaangażowania w dalszy rozwój tego narzędzia;

2) iteracyjny sposób rozwoju hurtowni danych, który w związku z stopniowym rozwojem hurtowni wymusza ciągle powtarzanie cyklu działań, polegających na specyfikacji potrzeb informacyjnych decydentów;

3) aktywny udział użytkowników w procesie rozwoju i modyfikacji hurtowni danych;

4) stosowanie metod prototypowania jako sposobu specyfikacji wymagań użytkowników. Jest to metoda angażowania użytkowników w proces specyfikacji ich potrzeb informacyjnych.

Etapem wprowadzającym do procesu budowy hurtowni danych powinno być opracowanie strategii budowy hurtowni danych. Zbierane informacje powinny dotyczyć: funkcji obsługiwanych i realizowanych przez hurtownię, przechowywanych danych, zastosowań źródłowych i analitycznych w ramach hurtowni, stosowanej technologii informatycznej, a więc sprzętu i oprogramowania, oraz przyszłej organizacji funkcjonowania.

Podczas gromadzenia danych o wymaganiach użytkowników hurtowni danych ważne jest osiągnięcie ogólnych celów analizy wymagań, tj. zrozumienie działań użytkowników hurtowni; ustalenie danych, które muszą się pojawić w projekcie początkowym hurtowni; zrozumienie sposobu wykorzystywania tych danych przez użytkowników; ustalenie danych wykorzystywanych wspólnie przez wielu użytkowników. Stworzony model musi spełniać wymagania funkcjonalne (potrzeby informacyjne) i niefunkcjonalne (parametry wydajności i dogodności użytkowania) decydentów i analityków informacji wobec modelu hurtowni danych.

Kolejną czynnością w procesie projektowania hurtowni danych jest realizacja projektu szczegółowego hurtowni. Najbardziej pracochłonnym działaniem w procesie tworzenia hurtowni danych jest projektowanie pozyskiwania, integracji i odwzorowania danych, które polega na: określeniu źródeł danych dla poszczególnych elementów danych zawartych w projekcie logicznym hurtowni; budowie algorytmów uzyskiwania i aktualizacji poszczególnych danych w hurtowni; określeniu zasad ujednociania danych źródłowych, jeżeli występują w więcej niż jednej bazie danych operacyjnych.

Naturalną cechą użytkowania hurtowni danych przez decydentów jest nieustanna zmiana potrzeb informacyjnych oraz pojawianie się nowych. Wdrożona i użytkowana hurtownia danych podlega więc ciągłej ewolucji. Ewolucja ta przebiega według podobnych etapów jak proces projektowania, tj. aktualizacja wymagań użytkowników i definiowanie nowych wymagań decyzyjnych (biznesowych); modyfikacja w modelu hurtowni danych; wykonanie projektu pozyskiwania i integracji danych oraz implementacja nowego rozwiązania⁶.

Elementy zapewniające osiągnięcie sukcesu

W tym podpunkcie koncentrować się należy na trzech kluczowych czynnikach, będących warunkiem powodzenia w każdej udanej implementacji hurtowni danych.

1. Koncentracja na biznesie, a nie na technologii

Skoncentrowanie się na biznesie oznacza wzięcie pod uwagę strategicznych dyrektyw i wyników finansowych firmy. Tworzenie hurtowni danych nie powinno być przedsięwzięciem technologicznym lub polegać tylko na zbieraniu danych z całej firmy w jednej wielkiej bazie danych. Musi być jasno określony cel biznesowy uzasadniający budowę hurtowni. Nie wystarczy po prostu stwierdzić, że „informacja jest strategiczna”, i dlatego firma potrzebuje hurtowni danych. Aby osiągnąć sukces, potrzeby muszą być dobrze zdefiniowane przy uwzględnieniu opłacalności inwestycji.

2. Szybkie, cykliczne reagowanie na potrzeby

Kluczem do sukcesu implementacji hurtowni danych jest szybkie, cykliczne reagowanie na potrzeby. W tym celu stosuje się metodę iteracyjnego rozwoju projektu. Należy przechodzić każdy etap wielokrotnie w serii iteracyjnych uściśleń odnoszących się do środowiska wspomagającego podejmowanie decyzji. Sprawą kluczową jest pozwolić na rozwój specyfikacji i implementacji w odpowiedzi na rozwój wymagań użytkowników końcowych,

⁶ *Inżynieria systemów informatycznych w E-gospodarce*, pod red. E. Kolbusz, W. Olejniczak, Polskie Wydawnictwo Ekonomiczne, Warszawa 2005, s. 222–230.

które pojawiają się podczas użytkowania systemu. Korekty i udoskonalenia w stosunku do poprzedniej iteracji mogą być wówczas dokonywane na podstawie reakcji użytkowników.

3. Użytkownicy końcowi w zespole implementującym

Bezpośredni udział użytkowników w zespole implementującym jest prawdopodobnie najważniejszym elementem w udanym tworzeniu hurtowni danych. Metoda iteracyjna działa dobrze jedynie, jeśli użytkownicy końcowi aktywnie biorą w niej udział. Taki udział musi polegać na czymś więcej niż tylko na uczestnictwie w posiedzeniach sprawozdawczych lub w comiesięcznych posiedzeniach grup wykonawczych. Użytkownicy końcowi powinni być bardzo mocno zaangażowani w podejmowania decyzji dotyczących rozwiązań projektowych⁷.

Zakończenie

W każdym z trafnych zastosowań hurtowni danych inwestycja zwraca z nawiązką nakłady, które firma ponosi na sprzęt, oprogramowanie i ludzi. Dobrze zaprojektowana hurtownia danych umożliwi więcej niż jedno jej zastosowanie, tak więc zwrot nakładów nie ogranicza się do jednego obszaru działania. Przykładów zastosowań hurtowni jest wiele, a możliwości jej wykorzystania do podejmowania decyzji sprawiają, że przedstawiona technologia staje się powszechnie stosowana. Zasadność jej wprowadzenia dla coraz większej liczby przedsiębiorstw jest oczywista. Obecnie świadomość, że informacja jest jednym z podstawowych czynników osiągania coraz większych zysków oraz przewagi konkurencyjnej, sprawia, że zarządzający nie mają problemów z podjęciem decyzji o wprowadzeniu do firmy analitycznych systemów Business Intelligence opartych na hurtowniach danych.

Literatura

- Imhoff C., Galleo N., Geiger J. G., *Mastering Data Warehouse Design – Relational and Dimensional Techniques*, Wiley Publishing, Inc., Indianapolis 2003.
- Inmon W. H., *Building the Data Warehouse, Fourth Edition*, Wiley Publishing, Inc., Indianapolis 2005.
- Inżynieria systemów informatycznych w E-gospodarce*, pod red. E. Kolbusz, W. Olejniczak, Polskie Wydawnictwo Ekonomiczne, Warszawa, 2005.
- Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P., *Hurtownie danych. Podstawy organizacji i funkcjonowania*, WSiP, Warszawa 2003.

⁷ V. Poe, P. Klauer, S. Brobst, op. cit., s. 98–103.

- Poe V., Klauer P., Brobst S., *Tworzenie hurtowni danych*, Wydawnictwa Naukowo-Techniczne, Warszawa 2000.
- Todman C., *Projektowanie hurtowni danych. Zarządzanie kontaktami z klientami (CRM)*, Wydawnictwa Naukowo-Techniczne, Warszawa 2003.

Construction and Functioning of a Data Warehouse in an Enterprise in the Context of the Effectiveness of the Decision Making Processes

Summary

In computerized enterprises, gathering huge amounts of data, traditional data processing would be highly ineffective and in turn fail to provide us with as interesting information, forecasts and relations, as Business Intelligence systems, for which Data Warehouses are the basis. The paper tries to answer a few questions regarding the nature Data Warehouse, the service it provides and its application. The author also discusses the stages of Data Warehouse construction and its influence on the effectiveness of the economic decision making processes within an enterprise.