

Dr Zygmunt Ryznar
Konsultant w Banku Przemysłowo-Handlowym (1998 r)

Hurtownie danych jako czynnik rozwoju (na przykładzie bankowości)

„Hurtownie danych to coś więcej niż gromadzenie danych, to właściwe i spójne odwzorowanie biznesu” (Hackathorn Richard)

Technologia hurtowni danych stała się nowością ostatnich lat¹ i występuje jako cecha charakterystyczna zaawansowanego etapu komputeryzacji, jaki następuje po automatyzacji obsługi operacyjnej, która w bankowości jak wiadomo polega na prowadzeniu rachunków klientów, dokonywaniu operacji na rynku pieniężnym i kapitałowym itp.

Przyczynami zapotrzebowania na technologię hurtowni danych są niedostatki stosowanych rozwiązań informatycznych oraz potrzeba posiadania globalnych informacji w skali firmy. Niedostatki polegają przede wszystkim na rozproszeniu danych w wielu systemach, ich wzajemnej niespójności, braku uporządkowania danych wg układów wymaganych do potrzeb zarządzania i niewydolności obsługi zadań w zakresie wielowymiarowych analiz informacji.

Niewydolność obsługi informacyjnej w operacyjnych systemach aplikacyjnych polega na tym, iż przetwarzanie złożonych zapytań i raportów powoduje znaczne spowolnienie operacji wykonywanych w trybie on-line. Czas reakcji systemu podczas obsługi okienkowej klientów może wówczas przedłużyć się z kilku sekund nawet do kilkunastu minut. Dotyczy to szczególnie żądań wymagających

¹ Technologia hurtowni danych pojawiła się jako reakcja na ograniczone możliwości obsługi złożonych zapytań odnoszących się do dużych wolumenów danych w środowisku tradycyjnie zaprojektowanych systemów zarządzania relacyjnymi bazami danych. Termin „data warehouse” został wprowadzony pod koniec lat 80-tych przez W.H.Inmona. Sformułował on takie cechy hurtowni danych jak: zorientowanie na problem, uwzględnienie wymiaru czasu, nieulotność, zintegrowanie ze środowiskiem użytkownika końcowego. Termin OLAP (On Line Analytical Processing) został utworzony w 1993 r. przez twórcę teorii relacyjnych baz danych E.F.Codda. Komitet d/s OLAPu powstał na przełomie 1994/1995 r. i w wyniku jego pracy pojawił się standard API dla dostępu i zarządzania wielowymiarowością danych. *OLAP oznacza przede wszystkim wielowymiarową analizę danych zainicjowaną przez końcowego użytkownika z jego stacji roboczej w trakcie oglądania danych na ekranie („on the fly”), obejmującą manipulację wymiarami (czyli „data-drilling”) oraz złożone mechanizmy raportowania i wizualizacji danych*

przejrzenia całych baz danych. Realizacja tego typu zadań, niezbędnych do uzyskania *globalnego spojrzenia na działalność bankową*, jest konieczna do sprawnego zarządzania finansami banku, które polega nie tylko i nie tyle na obsłudze poszczególnych rachunków klientów, ile na wyborze określonej strategii zdobywania kapitału oraz inwestowania, opartej na metodach pomiaru ryzyka bankowego i ukierunkowanej przez badania marketingowe.

W warunkach stosowania hurtowni danych, umożliwiającej rozdzielanie sfery informacji wykonawczych od sfery informacji decyzyjnych, nabierają „treści” (i odpowiedniej formy!) systemy wspomagania kierownictwa typu EIS i DSS. Skróty EIS ma „swoje lata” i bywa w związku z tym rozmaicie interpretowany. W wąskiej interpretacji EIS (*Executive Information System*) oznacza technologię obsługi zapytań i prezentacji informacji składowanych poza systemami obsługi operacyjnej, zaś DSS (*Decision Support System*) obejmuje metody modelowania i symulacji procesów decyzyjnych, w tym analizy typu „what if”.

Hurtownia danych jest jedną z metod scalania danych, rozproszonych po różnych aplikacjach i komputerach. Pozwala ona na odciążenie operacyjnego przetwarzania od utrzymywania informacji w różnych agregacjach i przekrojach oraz od ich maszynochłonnego przeszukiwania. W hurtowniach danych zwykle gromadzone są informacje w przekroju podmiotowym (klientowskim), walutowym i produktowym oraz czasowym (dzień, tydzień, dekada, miesiąc, kwartał, półrocze, rok, itp.).

Hurtownie danych pełnią więc ważną rolę integracyjną, szczególnie w niejednorodnym systemie bankowym, ale za ich pomocą nie można zlikwidować luk integracyjnych na poziomie operacyjnym, np. gdy specjalistą opiekującym się klientem potrzebna jest pozycja klienta a dilerowi aktualna pozycja waluty w banku. Hurtownie wyposażone w mechanizmy obsługi wielowymiarowych baz danych umożliwiają za to szybkie uzyskiwanie informacji -poprzez penetrację danych („data-drilling”)² na różnych poziomach analityczno-agregacyjnych oraz pozyskiwanie wiedzy w danych poprzez „data-mining”³.

² data-drilling oznacza penetrację, eksplorację danych (dosłownie „drażnienie danych”), obejmującą operacje uszczegóławiające (zwane też analizą wglębną, drażeniem w dół czyli „drill-down”), operacje agregujące (drażnienie w górę czyli „drill up”), analizę wieloprzekrojową („slicing & dicing”) itp.

³ Data mining jest technologią pozyskiwania wiedzy poprzez stosowanie modeli sieci neuronowych i algorytmów genetycznych, drzew decyzyjnych), technik statystycznych (np. regresji liniowych), modeli fraktalnych, algorytmów segmentacji (asocjacji, sekwencji, najbliższego sąsiedztwa) itp. W bankowości „data-mining” stosowane jest np.

Szybkie uzyskiwanie informacji w technologii hurtowni danych może być niewątpliwie poważnym *źródłem dochodowości banku* z tytułu zmniejszenia pracochłonności raportowania, lepszego zarządzania finansami banku (szybsze informacje o płynności), doboru lepszych i tańszych kanałów dystrybucyjnych dostosowanych do zachowania się klas (segmentów, profili) klientów, skrócenia czasu udzielania kredytów (np. dzięki automatycznej ocenie punktowej wniosku kredytowego), lepszego doboru marży w stosunku do kosztów transakcji, lepszego doboru klientów (np. bardziej dochodowych dla banku i o mniejszym stopniu ryzyka itp.

Hurtownia danych jest to scentralizowana nietransakcyjna baza danych przeznaczona do przechowywania informacji w długim horyzoncie czasowym globalnie w skali instytucji, w wielowymiarowych układach analitycznych i ukierunkowana na wyszukiwanie informacji bezpośrednio przez końcowych użytkowników.

Hurtownie danych stosowane są głównie do obsługi potrzeb zarządczych i sprawozdawczych, nie wymagających dostarczenia informacji w czasie rzeczywistym, aczkolwiek występują również przypadki ich użycia do wspomagania obsługi operacyjnej (np. do wykrywania nadużyć przy akceptacji kart płatniczych). Często w hurtowniach przechowywane są dane transakcyjne, a nie tylko zagregowane. Jest to niezbędne np. w przypadku „data-mining” oraz polepsza elastyczność gdyż daje możliwość tworzenia nowych agregatów danych.

Hurtownie danych „wpisują się” dobrze w ogólny trend przemian biznesowych, społecznych i technologicznych (patrz tabl.1). Przemiany te dotyczą przede wszystkim zachowania się banków na zmieniającym się rynku usług finansowych, a więc wzmocnienia i różnicowania działalności marketingowo-promocyjnej wobec narastającej konkurencyjności (nowi gracze na rynku i działalność parabankowa instytucji niebankowych) oraz zmian w zachowaniu klientów (różnorodność i wielokulturowość, wzrost wymagań klientów pod względem wygody, jakości i cen), prowadzeniu działalności inwestycyjnej i kredytowej w oparciu o wyrafinowane metody szacowania/obliczania ryzyka finansowego, wzmocnieniu pozycji wyjściowej banków poprzez konsolidacje organizacyjne i inne formy łączenia kapitału, itd.

Do realizacji tych przemian z reguły wymagane są zarówno gromadzone systematycznie przez dłuższy okres globalne informacje w skali banku i rynku,

do szacowania ryzyka w ocenie wniosków kredytowych, wykrywania oszustw, segmentacji klientów.

jak i odpowiednie narzędzia ich analizy. Właśnie hurtownie danych są przeznaczone do wykonywania tych zadań.

Tabl.1 Przemiany zachodzące w bankowości

stan dotychczasowy	zmiana
1. postawa reaktywna (oczekiwanie i reagowanie na napływające zdarzenia)	postawa proaktywna (uprzedzanie zjawisk niekorzystnych, generowanie wyprzedzające zdarzeń np. pobudzanie klientów do działania poprzez oferowanie nowych negocjowalnych produktów)
2. informacje izolowane w ramach poszczególnych aplikacji	zintegrowane informacje w skali firmy
3. ukierunkowanie na produkty	ukierunkowanie na rynek i klientów
4. marketing masowy (mass m.)	marketing ukierunkowany (target m., micromarkets) na segmenty klientów, marketing indywidualny (one-to-one marketing)
5. produkty definiowane a priori	produkty uzgadniane indywidualnie z klientem
6. obsługa okienkowa w oddziałach	elektroniczne kanały dystrybucji produktów, nowoczesna technologia po stronie klienta
7. klient tylko jako właściciel rachunków debetowych i kredytowych	zaspokajanie potrzeb klienta jako inwestora: (wspomaganie zarządzania przepływem pieniężnym klienta, wspomaganie zarządzania portfelem inwestycyjnym klienta)
8. raporty	prezentacje graficzne na ekranie uruchamiane przez końcowego użytkownika
9. aplikacje obsługi operacyjnej	hurtownie danych i systemy z inteligencją biznesową (stosowanie data mining)
10. analityk bankowy	knowledge worker korzystający z hurtowni danych
11. hierarchiczne struktury organizacyjne	dynamiczne poziome struktury oparte na współpracy funkcjonalnej

Źródło: opracowanie własne Autora.

Hurtownia w powyższym zestawieniu występuje zarówno jako czynnik samodzielny (poz.8,9,10) jak i dynamizujący większość pozostałych czynników rozwoju.

Na przykład jednym z podstawowych zastosowań hurtowni danych jest nowoczesny marketing, obejmujący wyprzedzające działania proaktywne (dotarcie do klienta nie tyle na podstawie odpowiedzi ankietowej, ile wskutek zdarzeń powodujących zmianę marketingowego statusu klienta (np. uaktywnienie na produkty typu: fundusze emerytalne, kredyty hipoteczne, kredyty studenckie, kasy

mieszkaniowe) np. osiągnięcie wieku emerytalnego, małżeństwo, narodziny dziecka, rozpoczęcie lub zakończenie studiów, zmiana adresu⁴.

Najważniejsze i najtrudniejsze problemy związane z tworzeniem hurtowni danych dotyczą definicji konkretnych celów biznesowych jakim mają one służyć, wyboru strategii budowy hurtowni danych (globalne hurtownie czy data marts⁵), ustalenia możliwie stabilnej struktury informacyjnej (w tym wymiarów) hurtowni danych, wyboru odpowiedniego oprogramowania narzędziowego oraz wdrożenia sprawnej technologii zasilania hurtowni danymi z wielu źródeł.

Podstawą warunkującą rozpoczęcie budowy hurtowni danych jest określenie stawianych przed nią celów biznesowych (np. utrzymanie klientów najbardziej dochodowych lub zdobywanie nowych klientów poprzez hurtownię marketingowo-klientowską, na podstawie której będzie można badać ich zachowanie i kształtowanie koszyków produktowych; lepszy dobór klientów i produktów pod kątem dochodowości, itp.) oraz zdefiniowanie wskaźników ekonomicznych dla obiektów biznesowych (tabl.2).

Tabl.2. Niektóre wskaźniki biznesowe w bankowości

Obiekt biznesowy	Wskaźniki-
rachunek	1.dochodowość 2.wartość dla banku 3.perspektywy
klient	1.wartość rankingowa dla banku 2.lojalność 3.dochodowość 4.cross-sell, koszyk produktów, rachunków ⁶ 5.perspektywy 6.koszyk transakcji 7.segmenty do których należy klient . 8. potrzeby klienta i stopień ich zaspokojenia 9.reklamacje klientów

⁴ Zmiana adresu dokonana w bazie klientów może oznaczać np. kupno domu, powiększenie stanu rodzinnego i traktować ją można jako oznakę zwiększonego zapotrzebowania na środki finansowe.

⁵ data marts (minihurtownie, podhurtownie, pseudohurtownie) są to podhurtownie wydzielone z globalnej hurtowni danych (zwane „dependent data marts”) albo wyciągi z operacyjnych baz danych (zwane „independent data marts” lub pseudohurtowniami). Data marts ukierunkowane są biznesowo na obsługę wydzielonych grup użytkowników (np. poszczególnych departamentów banku, zespołów problemowych).

⁶ ew. układ rachunków wewnątrz typów produktów (np. lokat)

	10. pozycja klienta w banku
produkt	1.dochodowość 2.miejsce w koszyku produktów 3.perspektywy 4.segment data-mining 5.nisze produktowe
oddział	1.dochodowość 2.miejsce w koszyku oddziałów 3.perspektywy
transakcja	1. dochody (opłaty prowizyjne i inne) 2. koszt transakcji
linie obsługi klienta (kanały dystrybucji produktów bankowych)	1.dochodowość 2.miejsce w koszyku linii dostaw produktów 3.perspektywy

Źródło: opracowanie własne Autora.

Projekty hurtowni danych powinny wynikać przede wszystkim z celów biznesowych a nie ambicji informatyków bankowych. Projektom tym można więc przypisać atrybut „business-driven”, a nie „information technology-driven”, co nie przekreśla użyteczności informatyków jako tych, którzy wskazują na możliwości (i ograniczenia) jakie stwarza technologia hurtowni danych, dobierają narzędzia prezentacji informacji odpowiednie dla grup użytkowników oraz administrują hurtowniami danych.

Hurtownia danych nie jest autonomiczną niezależną aplikacją, lecz przedsięwzięciem tworzącym globalną infrastrukturę informacyjną instytucji gromadzącą dane w długim horyzoncie czasowym (5-7 lat) do obsługi celów biznesowych.

Hurtowni danych nie można zbudować bez udziału kierownictwa banku, którego głównym zadaniem jest tutaj pełnienie roli sponsora przedsięwzięcia i ukierunkowanie celów biznesowych przez zdefiniowanie strategii finansowej banku.

Jednym z podstawowych wymagań (i trudności) przy definiowaniu potrzeb informacyjnych pod kątem hurtowni danych jest konieczność oderwania się od myślenia kategoriami sztywnych raportów na rzecz wielowymiarowej analizy informacji oraz umiejętność przechodzenia od stwierdzenia faktów do szukania przyczyn.

Zygmunt Ryznar