

Hanna Borzęcka

Wojskowa Akademia Techniczna w Warszawie

## Wielowymiarowe struktury danych OLAP odpowiedzią na współczesne problemy informacyjne zarządzania

*„W erze informacji wszyscy są w stanie gromadzić olbrzymie ilości danych o konsumentach, klientach czy rynku. Wygrają nie ci, których bazy będą największe, lecz ci, których narzędzia będą umożliwiały wszechstronną ich analizę.”*  
Bill Gates: „Biznes szybki jak myśl”

### Streszczenie

Publikacja zawiera odpowiedź na pytanie, czym jest hurtownia danych i jakie są korzyści z jej wdrożenia. Definiuje dzisiejsze potrzeby informacyjne zarządzania przedsiębiorstwem oraz określa rolę hurtowni danych w zaspakajaniu tych potrzeb.

W artykule omówiono koncepcję wielowymiarowych struktur danych oraz przedstawiono, zbudowany wg koncepcji autorki, sformalizowany wielowymiarowy model danych podstawowej struktury hurtowni danych, czyli kostki OLAP (ang. On-Line Analytical Processing). Przedstawiono również wyzwania stojące aktualnie przed współczesnymi hurtowniami danych w zakresie ich eksploracji i odkrywania z nich wiedzy.

Publikacja zawiera także propozycję innowacji i przewiduje powstanie nowej generacji hurtowni danych rozwiązującej problemy zarządzania przedsiębiorstwem w warunkach niepewności.

### Wprowadzenie

We współczesnym, dynamicznie zmieniającym się świecie, pozycja przedsiębiorstwa jako jednostki funkcjonującej w zmieniającym się otoczeniu jest wysoce zagrożona. W erze globalizacji rosnąca konkurencja między firmami wymusza na nich nieustanną obserwację i analizę zachodzących zmian oraz szybką reak-

cję wewnątrz i na zewnątrz organizacji w celu utrzymania pozycji rynkowej. Cechy dzisiejszego otoczenia gospodarczego to globalizacja, elektronizacja, wirtualizacja i „usieciowienie” rynku, konkurencja w każdym miejscu, w każdej chwili, z każdej strony, a także gwałtowny wzrost liczby urzędzeń i systemów teleinformatycznych. Zdarzenia i zjawiska gospodarcze rozprzestrzeniają się dosłownie z prędkością światła. Menedżerowie zasypywani są ogromną ilością danych w większości dziedzin. Przedsiębiorstwa funkcjonują w cyberprzestrzeniach, w których czas przesyłania informacji i czas podejmowania decyzji jest bardzo krótki. Powstaje nowa forma rynku – rynek elektroniczny, oraz nowa forma organizacji – organizacja wirtualna, a także nowa forma społeczeństwa – globalne społeczeństwo informacyjne.

Zachodzące zmiany powodują konieczność zmian w zarządzaniu. Kierunek zmian to konieczność zastosowania technologii informacyjnych w organizacjach.

## 1. Informacyjne i integracyjne problemy zarządzania

Jak twierdzi J. Kelly, „integracja biznesu jest podstawowym celem integracji systemów przetwarzania informacji”. Wkroczyliśmy zatem w okres, kiedy informacja stała się podstawą funkcjonowania przedsiębiorstw, wszystkich szczebli administracji oraz życia jednostek. Przedsiębiorstwa nie są już usatysfakcjonowane samym gromadzeniem ogromnej ilości przeprowadzonych transakcji. Potrzebne jest właściwe przetworzenie i segregacja tych informacji, np. aby sprawdzić, czy kluczowe wskaźniki ekonomiczne jednostki organizacyjnej poprawiły się.

Firmy osiągają zysk tylko wtedy, gdy są w stanie analizować zmieniające się trendy w popycie i równie szybko się do nich dostosowywać. Informacja powinna być przetworzona i przedstawiona w sposób jak najbardziej przydatny dla użytkownika, a dostarczona w niezbędnym dla niego miejscu i czasie. Przedsiębiorstwom potrzeba coraz więcej informacji przetworzonej oraz ukierunkowanej na konkretne problemy i obszary analizy informacji w zakresie:

- wielkości sprzedaży w podziale na klientów, produkty, obszary sprzedaży, przedstawicieli handlowych, sklepy, oddziały, jednostki powiązane, kanały dystrybucji itd.,
- kosztów, w podziale na miejsca ich powstania (MPK), rodzaje (koszty bezpośrednie i pośrednie, koszty stałe i zmienne), centra i nośniki kosztów, koszty planowane, koszty działalności podstawowej, koszty działalności pomocniczej, koszty zarządu, koszty procesu sprzedaży, koszty zakupu itp.,
- rabatów i promocji,
- „wąskich gardeł”,
- wielkości zysku, w podziale na wszelkie interesujące jednostkę gospodarującą przekroje,
- rotacji produktów,
- sezonowości sprzedaży,
- bieżącej analizy trendów,

- kalkulacji marż, w podziale na kanały hurtowe, regiony geograficzne, klientów, rodzaje produktów itp.,
- przepływów pieniężnych i budżetów,
- zamówień klientów,
- wskaźników ekonomiczno-finansowych.

Konieczność wieloprzekrojowych analiz w celu ujęcia działalności przedsiębiorstwa we wszystkich jakościowo różnych aspektach jego działalności stwarza konieczność posługiwania się modelami przedsiębiorstw (ang. *corporate models*). Brak modelu przedsiębiorstw uniemożliwia bowiem kompleksowe, wewnętrznie zgodne analizowanie procesów gospodarczych.

W zarządzaniu przedsiębiorstwem niezwykle istotne są problemy integracji, ponieważ problemy informacyjne w organizacjach wynikają w dużej mierze z problemów integracji środowiska zarządzania. Integracja to połączenie niejednorodnych składników w całość tak, że współdziałając w ramach tej całości, wzmagają swoją skuteczność. Należy też podkreślić synergiczny efekt integracji.

Niezbędna jest integracja zarządzania organizacją i systemu informacji. Integracja ta powinna być zorientowana na wspomaganie decyzji. Rozwiązaniem problemów informacyjnych i integracyjnych w organizacji jest informacyjny system zarządzania. Jest to wielopoziomowa, polihierarchiczna sieć powiązań, której węzłami są wszystkie elementy organizacji i jej otoczenia, biorące udział w wymianie informacji. Integracja w informacyjnym systemie zarządzania powinna być oparta na inżynierii systemów. Kluczowy jest zintegrowany system wspomaganie decyzji.

## 2. Informatyka narzędziem zarządzania organizacją XXI wieku

Technologia informacyjna jest podstawowym wsparciem zarządzania XXI wieku. Wiele przedsiębiorstw aktywnie poszukuje nowych technologii, które pomogą im stać się bardziej dochodowymi i konkurencyjnymi. Zdobywanie przewagi nad konkurencją wymaga od przedsiębiorstw przyspieszenia procesu podejmowania decyzji tak, aby szybko mogły reagować na zmiany. Kluczowym w przyspieszaniu procesu podejmowania decyzji jest posiadanie w odpowiednim czasie właściwej i łatwo dostępnej informacji. To właśnie cybernetycy pierwsi zwrócili uwagę na fakt, że wszystkie procesy można pojmować w kategoriach informacyjnych. Już w 1948 r., C.E. Shannon opublikował pracę pt. „*A Mathematical Theory of Communication*”, w której stworzył fundamentalne podstawy teorii informacji, i wraz z Wienerem wykazał, że informacja jest jednym z podstawowych atrybutów materii – obok masy, czasu, energii i przestrzeni.

W 1952 r. A. Turing opublikował artykuł pt. „*Computing Machinery and Intelligence*”, w którym postawił pytanie, czy maszyna może myśleć. Turing sformułował podstawy sztucznej inteligencji, które zainicjowały rozwiązywanie problemów badań cybernetycznych. J. von Neumann w swej pracy z 1958 r. pt. „*The Komputer and the Brain*” wykazał związki między budową i funkcjonowaniem ludzkiego mózgu i komputera.

W latach 70. ubiegłego wieku narodziła się informatyka, bez której aktualnie trudno wyobrazić sobie nowoczesne zarządzanie organizacją. Technologie informacyjne mają obecnie kluczowe znaczenie dla gospodarki. Gwałtowny, dynamiczny rozwój technologii informacyjnych stwarza nowe możliwości dla przedsiębiorstw, daje szanse na sukces i konkurencyjność w stosunku do innych. Wiele przedsiębiorstw decyduje się jednak na nowe nakłady na technologie informacyjne pod wpływem nacisku, mody i presji otoczenia oraz reklamy, bez długofalowej wizji i świadomości wszelkich wynikających z tego konsekwencji. Firmy wydają znaczne środki na nowe technologie bez uzyskania zadowalających rezultatów. Aby tego uniknąć potrzebna jest nowa, długofalowa strategia informacyjna dla przedsiębiorstw i następnie zbudowana na jej podstawie strategia informatyczna.

### **3. Hurtownie danych – podstawą nowoczesnej strategii informacyjnej w zarządzaniu**

P. Drucker, w swojej pozycji pt. „Zarządzanie w XXI wieku” napisał, iż: „Zarządzanie przedsiębiorstwem oznacza zarządzanie przyszłością, a zarządzanie przyszłością oznacza zarządzanie informacją”.

Strategiczne zarządzanie jest imperatywem biznesu. Wykorzystanie współczesnej technologii informacyjnej wymaga również strategicznego planowania ze względu na zakres i skalę ich stosowania, duże koszty i konieczności zastosowania zmian organizacyjnych i czasu niezbędnego do wzajemnego dostosowania potrzeb. Technologie informacyjne same w sobie nie dają już przewagi. Liczy się to, kto jak potrafi je wykorzystać. Dzisiejsze potrzeby zarządzania zaspokajają w zakresie wsparcia procesów podejmowania decyzji technologia informacyjna hurtowni danych, której celem jest dostarczanie właściwych informacji właściwym ludziom, we właściwym czasie, a jednocześnie po niskim koszcie. Hurtownie danych zostały zaprojektowane dla osób podejmujących strategiczne decyzje w przedsiębiorstwie, ale służą także do monitorowania efektów podjętych decyzji.

Hurtownie danych stały się architektoniczną odpowiedzią na chaos najróżniejszych systemów, platform, środowisk i niezgodnych danych. Zadaniem ich jest dostarczanie informacji ludziom podejmującym decyzje w firmach tak, aby mogli oni zanalizować sytuację i podjąć decyzję. Dzięki technologii wielowymiarowej struktury danych można budować aplikacje umożliwiające analizę wszystkich obszarów działalności przedsiębiorstwa – finansów, sprzedaży, klientów, marketingu, dystrybucji itd. Pozwala to nie tylko skutecznie mierzyć biznes, ale również optymalnie go planować.

Najważniejsze cechy hurtowni danych to:

**1. Możliwość pozyskiwania danych z różnych, niezależnych źródeł danych.** Informacje potrzebne do analiz mogą pochodzić z wielu różnych aplikacji, zbudowanych na różnych platformach informatycznych. Technologie hurtow-

ni danych używają narzędzi ETL (ang. *Extract-Transformation-Loading*) do procesu zasilania hurtowni danych z różnych systemów źródłowych. Jest to czyszczenie danych, porządkowanie, konsolidacja i przekształcanie do wspólnego formatu, ponieważ hurtownia danych zasilana jest na ogół z wielu, rozproszonych i różnorodnych źródeł danych.

**2. Wolumen danych.** Hurtownia danych ma możliwość przechowywania ogromnych baz danych (rzędu 400 Gigabajtów).

**3. Narzędzia analityczne OLAP.** Umożliwiają wykonywanie analiz wielowymiarowych. Jest to możliwość tworzenia i analitycznego przetwarzania modelu wielowymiarowego danych – zorientowanego na potrzeby biznesowe. Pojęcie analizy wielowymiarowej stało się podstawą metody definiowania hurtowni danych. Narzędzia OLAP wyposażone są w język MDX (ang. *MultiDimensional Expressions*), który pozwala na tworzenie, zarządzanie i analizę danych zdefiniowanych jako struktury wielowymiarowe. Jest to rozszerzenie standardowego języka SQL. Dzięki nowym funkcjom możliwe są np. agregacje (operacja *Cube*) i działania na hierarchiach (operacja *Rollup*), które umożliwiają tzw. zwijanie i rozwijanie informacji zawartych w kostkach OLAP. Głównym celem hurtowni danych jest zwiększenie efektywności i wydajności procesu podejmowania decyzji w przedsiębiorstwach. Hurtownię danych przyjęto definiować wg Inmona (1996) jako zbiór zintegrowanych, nieulotnych, ukierunkowanych baz danych, wykorzystywanych w systemach wspomagania decyzji.

**4. Zintegrowanie,** które oznacza, że dane są jednolite. Przed umieszczeniem danych z wielu różnych, niezależnych źródeł w hurtowni danych niezbędny jest proces integracji danych co oznacza, że będą one ujednoczone, np. daty będą przechowywane zawsze w tym samym formacie, niezależnie od formatów dat z systemów źródłowych. Technologia hurtowni danych jest niezależna od platformy sprzętowej, systemu operacyjnego, czy baz danych systemów źródłowych.

**5. Nieulotność,** co oznacza, że dane raz umieszczone w hurtowni, zazwyczaj pozostają niezmiennione. Oznacza to, że zapytanie zwróci ten sam wynik, niezależnie od tego jak często jest wykonywane.

**6. Zorientowanie na temat.** Oznacza to, że zbierane dane dotyczą pewnego tematu (np. sprzedaży), a nie działań wynikających z obsługi procesów biznesowych.

**7. Określoność w czasie.** Dane w hurtowni są poprawne tylko dla określonego okresu czasu.

Oddzielna, stała baza danych jest podstawową składową architekturą hurtowni danych, ponieważ są podstawowe różnice w działaniu systemów operacyjnych i systemów wspomagania decyzji. Jedna baza danych nie może być bowiem funkcjonalna, zarówno w przetwarzaniu operacyjnym, jak i wspomagającym podejmowanie decyzji. W obu tych przypadkach wymagane są bazy danych różnych rodzajów. Między systemami operacyjnymi i hurtowniami danych występuje konflikt spowodowany tym, że systemy operacyjne nie są zorientowane na temat, lecz przygotowane do zaspokajania wymagań operacyjnych. Systemy operacyjne podlegają ciągłym zmianom. Operacyjne bazy danych są przeznaczone do przechowy-

wania, aktualizowania i szybkiego składania raportów na temat aktualnych danych biznesowych. Systemy wspomagania decyzji mają natomiast za zadanie umożliwić prosty dostęp do stałych, historycznych i analitycznych informacji.

Spór o zasoby dla systemów informatycznych w dużych przedsiębiorstwach jest zawsze poważnym zagadnieniem. Systemy wspomagania decyzji znane są z tego, że pochłaniają ogromne ilości środków. Realizacja skomplikowanego zapytania, które najczęściej polega na przeglądaniu, łączeniu i gromadzeniu wiele milionów wierszy, stwarza ogromne zapotrzebowanie na pamięć, czas pracy procesora itp. Gdy system wspomagania decyzji dzieli środowisko z systemami operacyjnymi, często kończy się to katastrofą. Za pierwszym pojawieniem się skomplikowanego zapytania do hurtowni danych podczas operacyjnego zapełniania bazy danych dużą ilością danych pojawia się konflikt w dostępie do zasobów. Problem zostanie rozwiązany poprzez przerwanie zapytania z systemu wspomagania decyzji, gdyż transakcje operacyjne są prawie zawsze traktowane jako ważniejsze od zadań wspomagania decyzji. Obsługa klientów ma pierwszeństwo.

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. Hurtownia danych to technologia informacyjna służąca przekształcaniu dużych wolumenów danych w informację, a następnie w wiedzę.

Głównym celem jest uzyskanie integralności danych i integracji na poziomie całego przedsiębiorstwa. Technologia ta jest niezależna od platform sprzętowych, systemów operacyjnych i baz danych. Łączy dane z wielu heterogenicznych środowisk. Dane są czyszczone i normalizowane, a następnie organizowane w celu dostarczenia informacji o głównych podmiotach zainteresowania: klientach, produktach, dostawcach, kosztach, sprzedaży. Hurtownia danych skupia się na dostarczaniu informacji o podmiotach, a nie na dziennych czynnościach związanych z podstawową działalnością firmy. Dostarcza informacji o podmiotach, które uczestniczą w procesie podejmowania decyzji. Dane pochodzące z różnych systemów są konsolidowane, agregowane i podsumowywane. Informacje potrzebne do analiz, które pochodzą z wielu różnych systemów informatycznych, zarówno wewnętrznych, jak i zewnętrznych są gromadzone w jednym miejscu. Korzystając z tej technologii można analizować wszystkie obszary działalności przedsiębiorstwa.

## 5. Struktura OLAP

Odpowiedzią na dzisiejsze problemy zarządzania jest technologia wielowymiarowej struktury danych. Jak uczy praktyka, często nadmiar informacji może być równie niebezpieczny, jak jej brak. Dlatego bardzo ważne jest właściwe segregowanie i agregowanie zgromadzonych danych oraz właściwa ich wizualizacja.

Dzięki technologii wielowymiarowej struktury danych można budować aplikacje, dzięki którym możliwa jest analiza wszystkich obszarów działalności

przedsiębiorstwa – finansów, sprzedaży, klientów, regionów, kanałów dystrybucji, kosztów, zysku, budżetów, marketingu, zapasów, rentowności w podziale na poszczególnych kontrahentów, przedstawicieli, obszary sprzedaży itd. Systemy OLAP są nową i coraz bardziej popularną oraz ważną klasą systemów informatycznych, zajmujących się zarządzaniem i analizą danych modelowanych w struktury wielowymiarowe. Najczęściej odbywa się to w środowisku hurtowni danych wyposażonych w narzędzia bezpośredniego przetwarzania analitycznego.

W codziennej działalności przedsiębiorstwa są wykorzystywane **systemy do przetwarzania operacyjnego OLTP** (ang. *On-Line Transaction Processing*), które obsługują aktualizujące dane na bieżąco. Systemy te mają przede wszystkim wspomagać funkcjonowanie przedsiębiorstwa przez obsługę transakcji. Natomiast systemy analityczne to takie, które przetwarzając informacje wspomagają procesy podejmowania decyzji. Dane wykorzystywane podczas przetwarzania analitycznego są historyczne (najczęściej projektuje się stałą bazę danych – tylko do odczytu). Użytkownicy tej bazy mogą przeprowadzać analizę trendów i wzorców danych na podstawie dużej ich liczby. Odrębna technologicznie, analityczna baza danych wykorzystywana jako podstawa systemu wspomagania decyzji to hurtownia danych. Wyposażona w narzędzia OLAP zapewnia dostęp do informacji i przetwarzanie analityczne.

Po raz pierwszy terminu OLAP użył Codd w 1993 r. w pracy pt. „*Providing OLAP to User-Analysts: An IT Mandate*”. Codd sformułował 12 reguł dla OLAP. Najważniejsze z nich to:

- wielowymiarowy model koncepcyjny danych (łatwy w zastosowaniu),
- architektura klient-serwer,
- wewnętrzna wymiarowość (każdy wymiar w bazie danych musi mieć odpowiednik w strukturze i możliwościach działania),
- nieograniczoność międzywymiarowych operacji (obsługa hierarchii wymiarów i agregacji),
- operowanie danymi wielowymiarowymi ( obracanie, zwijanie, rozwijanie),
- nieograniczoność wymiarów i poziomów agregacji.

Systemy OLAP są samodzielną klasą systemów informatycznych wyposażonych w narzędzia do tworzenia hurtowni danych, pobierania danych z systemów operacyjnych, zarządzania hurtownią danych oraz w narzędzia analityczne. Aplikacja OLAP jest zwykle instalowana na specjalnym serwerze OLAP, który stosuje wielowymiarowe struktury do przechowywania danych i relacji. OLAP umożliwia wielowymiarową analizę danych przechowywanych w wielowymiarowych strukturach w postaci tzw. kostek OLAP.

Jednym ze sposobów stosowanych podczas projektowania hurtowni danych jest tworzenie i implementacja modelu wielowymiarowego. Dało to początek analizie wielowymiarowej. Podstawowe pojęcia, jakie wyróżnia się w modelu wielowymiarowym OLAP, to:

- **Tabela faktów**, czyli podstawowa tabela, w której zawarte są **miary** lub inaczej wielkości, których analizą jesteśmy zainteresowani. Jest to relacyjna tabela danych pamiętana w bazie danych (hurtowni danych) i zawiera-

jąca informacje, które mają podlegać analizie w systemie OLAP, np. koszt, sprzedaż itp.

- **Miary**, definiowane jako te cechy, których wartości są istotne dla użytkownika. Wyznaczane są na podstawie tabeli faktów. Wartości ich wyznaczone są w wyniku pomiarów w przestrzeni wielowymiarowej. Wartości miar mogą być rozważane na różnych poziomach agregacji, np. miesiąc, kwartał, rok (wymiar czasu) lub produkt, grupa produktów, marka produktów (wymiar produktu). Uzyskanie zagregowanych wartości miar odbywa się za pomocą odpowiedniej funkcji agregującej stosowanej do wskazanej kolumny tabeli faktów albo przez przeliczenie wartości innych miar wcześniej zagregowanych.
- **Funkcje agregujące**, które określają, w jaki sposób z danych elementarnych tworzone są dane zagregowane.
- **Wymiary**, które określają przestrzeń, w której analizowane są miary.
- **Elementy (człony) wymiarów**, służące do określania punktów w przestrzeni wielowymiarowej.
- **Komórki**, które zawierają wartości miar w określonym punkcie przestrzeni wielowymiarowej.
- **Kostka**, czyli zbiór wszystkich możliwych komórek.

Dzięki technologiom OLAP możliwe jest rozliczanie działalności przedsiębiorstwa w ujęciu wieloprzekrojowym (jednocześnie wg komórek organizacyjnych, MPK, segmentów rynku, regionów geograficznych, kanałów dystrybucji), wielowymiarowym oraz wielookresowym.

Technologia OLAP pozwala użytkownikom przeprowadzać kompleksowe analizy danych poprzez szybki dostęp do wielowymiarowych widoków przedsiębiorstwa. OLAP pozwala odpowiedzieć nie tylko na pytania: „kto?”, „co?”, „kiedy?”, ale też na pytania: „co jeśli?”, „dlaczego?”. Aplikacje OLAP pozwalają na prognozowanie przyszłości na podstawie danych historycznych. Analiza wielowymiarowa pozwala na odkrywanie zależności, które nie mogą być zauważone bezpośrednio z surowych danych. Wielowymiarowość jako cecha OLAP umożliwia dokonywanie analiz poprzez elastyczny dostęp do informacji. Ważne jest, aby aplikacje OLAP dostarczały decydom informacji, których potrzebują oni do podejmowania efektywnych decyzji.

## 6. Wielowymiarowy model danych OLAP

Przedstawiony ogólny opis daje się opisać formalnym modelem matematycznym. Te same pojęcia, co wcześniej użyte na potrzeby opisu modelu należy sformułować następująco:

**Wymiary** – niezależne, rozłączne zbiory elementów danego typu (niekoniecznie liczbowego). Wymiary definiują przestrzeń, w której dokonujemy analizy. Każdy wymiar organizowany jest poprzez określenie: hierarchii (opcjonalnie), pozio-



mów (opcjonalnie) i elementów lub członów (obligatoryjnie). Ilość i rodzaj wybranych wymiarów oraz związane z nimi hierarchie, poziomy i atrybuty zależą od tego, jakiego rodzaju analizę przeprowadzamy, i z jakiego punktu widzenia rozpatrujemy dane zjawiska (miary).

**[Zbiór wymiarów w kostce OLAP:]**

$(D_1, D_2, \dots, D_i, \dots, D_N)$  – gdzie:  $D_i$  – zbiór opisujący i-ty wymiar,  $N$  – ilość wymiarów w kostce OLAP.

**[Zbiór elementów wymiaru  $D_i$ :]**

$D_i = \{d_i^1, d_i^2, \dots, d_i^j, \dots, d_i^k\}$  – gdzie:  $d_i^j$  – j-ty element i-tego wymiaru;  $k$  – ilość elementów w wymiarze i-tym.

W kostce  $C$  o  $N$  wymiarach zbiór  $D$  określamy jako iloczyn kartezjański wymiarów:  $D = X D_n = D_1 \times D_2 \times \dots \times D_i \dots \times D_N$ , gdzie  $D$  – jest zbiorem punktów przestrzeni  $N$  – wymiarowej kostki OLAP.

**[Hierarchie i poziomy wymiarów:]**

Zbiór elementów wymiaru może być podzielony na rozłączne podzbiory – poziomy, uporządkowane hierarchicznie zgodnie ze stopniem ogólności ich znaczeń:

$L^h \leq L^{h-1} \leq \dots \leq L^1 \leq L^0$ , gdzie:  $L^0$  – poziom najwyższy,  $L^h$  – poziom najniższy,  $h$  – głębokość poziomów i np. dla  $h = 0$  wymiar posiada tylko jeden poziom dla swoich członów, co oznacza, że nie mogą być utworzone hierarchie dla tego wymiaru.

Ułożone hierarchicznie poziomy tworzą hierarchię wymiaru:

$H_{ij} = \{L^{0_{ij}} \geq L^{1_{ij}} \geq \dots \geq L^{h_{ij}}\}$  – gdzie:  $H_{ij}$  – i-ta hierarchia j-tego wymiaru,  $i = 0$  – brak hierarchii dla j-tego wymiaru:  $L^0_j = \{D_j\}$  czyli jest tylko jeden poziom, który zawiera wszystkie elementy wymiaru;  $h$  – głębokość i-tej hierarchii j-tego wymiaru.

**[Atrybuty wymiarów:]**

Z każdym poziomem wymiaru może być związany zbiór Atrybutów.

Zbiory atrybutów dla wymiaru określa wektor m-wymiarowy:

$A = (A_1, A_2, \dots, A_i, \dots, A_m)$  – gdzie:  $A_i$  – i-ta zmienna określająca zbiór wartości i-tego atrybutu,  $m$  – ilość atrybutów dla wymiaru.

Wektor atrybutów  $A$  przyjmuje skończoną liczbę wektorów  $a = (a_1, a_2, \dots, a_i, \dots, a_m)$ . Ilość wektorów  $a$  jest równoznaczna z ilością elementów wymiaru, dla którego utworzono wektor atrybutów  $A$ . Każdemu j-temu elementowi wymiaru jest przyporządkowany wektor  $a^j$ , którego składowe są wartościami poszczególnych atrybutów dla danego elementu:

$a_j = (a_{j1}, a_{j2}, \dots, a_{ji}, \dots, a_{jm})$  – gdzie:  $a^j$  – wektor wartości atrybutów dla j-tego elementu wymiaru,  $a_{ij}$  – wartość i-tego atrybutu dla j-tego elementu,  $j = 1, 2, \dots, k$ ,  $k$  – ilość elementów wymiaru.

Wektor atrybutów  $A$  może być związany z każdym poziomem każdego wymiaru.

**[Miary:]**

Miary definiowane są jako te cechy, których wartości są istotne dla użytkownika w procesie analizy. Są wyliczane na podstawie wskazanej tabeli faktów, w wyni-

ku pomiarów w przestrzeni wielowymiarowej dla wszystkich wartości iloczynu kartezjańskiego wymiarów. Wartości miar są określane dla wszystkich poziomów agregacji wynikających z hierarchii wymiarów (np. rok, kwartał, miesiąc). Uzyskanie wartości miar odbywa się za pomocą funkcji agregującej zastosowanej do wskazanej kolumny z tabeli faktów lub poprzez przeliczenie wartości innych miar wcześniej zagregowanych. Miary mają wyłącznie wartości liczbowe.

**Definicja Miary.** W każdej kostce możemy definiować dowolną ilość miar:

$$\begin{array}{c} \mathbf{N} \\ \mathbf{F}_1: \mathbf{X} \mathbf{D}_n \rightarrow \mathbf{M}_1 \\ n=1 \\ \mathbf{N} \\ \mathbf{F}_2: \mathbf{X} \mathbf{D}_n \rightarrow \mathbf{M}_2 \\ n=1 \\ \dots \quad \dots \\ \mathbf{N} \\ \mathbf{F}_i: \mathbf{X} \mathbf{D}_n \rightarrow \mathbf{M}_i \\ n=1 \\ \dots \quad \dots \\ \mathbf{N} \\ \mathbf{F}_m: \mathbf{X} \mathbf{D}_n \rightarrow \mathbf{M}_m \\ n=1 \end{array}$$

gdzie:

$\mathbf{F}_i$  – funkcja przekształcająca iloczyn kartezjański zbiorów  $\mathbf{D}_1 \times \mathbf{D}_2 \times \dots \times \mathbf{D}_i \dots \times \mathbf{D}_N$  wymiarów w zbiór  $\mathbf{M}_i$  miary i-tej.

Zbiór  $\mathbf{F}$  wszystkich funkcji przekształcających iloczyn kartezjański  $\mathbf{D}_1 \times \mathbf{D}_2 \times \dots \times \mathbf{D}_i \dots \times \mathbf{D}_N$  zbiorów wymiarów w zbiór wszystkich miar:

$$\mathbf{F} = \langle \mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_m \rangle \mathbf{X} \mathbf{D}_n \rightarrow \mathbf{M} = \langle \mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_i, \dots, \mathbf{M}_m \rangle$$

#### [Wielowymiarowy model danych OLAP:]

Kostka OLAP to uporządkowana trójka:

$$\langle \{\mathbf{D}_n\}_{n \in \mathbf{N}}, \{\mathbf{M}_k\}_{k \in \mathbf{K}}, \mathbf{F} \rangle$$

gdzie:

$\{\mathbf{D}_n\}_{n \in \mathbf{N}}$  – rodzina zbiorów wymiarów

$\mathbf{N}$  – ilość wymiarów

$\mathbf{D}$  – uniwersum, zbiór wszystkich obiektów (stanów) systemu

$$\mathbf{D} = \mathbf{X} \mathbf{D}_n = \mathbf{D}_1 \times \mathbf{D}_2 \times \dots \times \mathbf{D}_i \dots \times \mathbf{D}_n$$

$\{\mathbf{M}_k\}_{k \in \mathbf{K}}$  – rodzina zbiorów miar,  $\mathbf{K}$  – ilość miar,  $\mathbf{F}$  – zbiór funkcji takich, że



## 8. Nowa generacja hurtowni danych

Zjawiska i procesy ekonomiczne ze względu na ich różnorodność i złożoność a jednocześnie konieczność poznania i zrozumienia mechanizmów stojących u ich podstaw, są przedmiotem intensywnych badań, wykorzystujących różnorodne narzędzia formalne. W związku z tym stanowią one interesujący obszar badawczy dla praktycznej weryfikacji przydatności analizy danych. W naukach ekonomicznych dużo jest nazw nieostrych, zarówno w języku potocznym, jak i w języku fachowym, np. ubóstwo, bogactwo, kryzys, inflacja, stabilizacja, wzrost gospodarczy, równowaga rynkowa itp. Pojęć nieostrych, z których nie chce zrezygnować ekonomia nie da się zastąpić liczbami. Potrzebny jest jakiś inny sposób pozwalający zachować nieostrość, ale zapewniający równocześnie jednoznacznie rozumienie znaczenia nazw nieostrych.

Matematyka klasyczna jest nieadekwatna do problemów społeczno-ekonomicznych, z powodu występowania w nich nieostrych nazw oraz z powodu specyficznego charakteru prawidłowości kształtowania się zjawisk społeczno-ekonomicznych. Jeden z możliwych sposobów matematycznego modelowania pojęć nieostrych polega na wykorzystaniu pojęcia zbioru rozmytego. Największy rozgłos w zakresie modelowania nieostrości zyskała propozycja L.A. Zadeha, który odrzucił dwuwartościową logikę oraz Boolowską teorię zbiorów, które nie są w stanie rozwiązać wielu sprzeczności i niejednoznaczności w modelowaniu i przetwarzaniu rzeczywistych danych.

Decyzje w systemach ekonomicznych charakteryzują się złożonością, zmiennością w czasie i noszą niedeterministyczny charakter, utrudniający znacznie podejmowanie decyzji. Proces podejmowania decyzji w tego typu systemach opiera się na niepełnej i rozmytej informacji. Dzięki teorii L.A. Zadeha, może nastąpić złagodzenie żądania precyzji modelu w kierunku rozmytej reprezentacji zmiennych i parametrów. W tym celu konieczne zastąpienie binarnego widzenia świata, rozmytą reprezentacją danych oraz modelowaniem i wnioskowaniem rozmytym. Zbiory rozmyte mogą się przyczynić do istotnego postępu w matematyzacji nauk społeczno-ekonomicznych.

Inne nowe podejście do analizy danych, oparte jest na podanej przez Pawlaka (1982) teorii zbiorów przybliżonych, które umożliwia zastosowanie teorii zbiorów przybliżonych do analizy danych ekonomicznych.

Aktualnie brak efektywnych algorytmów i programów komputerowych wspomagających podejmowanie decyzji w zadaniach rozmytych. Nowy kierunek to przezwyciężanie wad tradycyjnych algorytmów komputerowych i opisywanie rzeczywistości w sposób naśladowujący rozumowanie człowieka. Szczególne znaczenie ma rozwój i powstanie nowej generacji hurtowni danych w oparciu o niedeterministyczne modele matematyczne, tzn. probabilistyczne (stochastyczne), strategiczne, rozmyte i przybliżone.

Szczególne znaczenie mają modele rozmyte, które znakomicie opisują nieostrości i nieokreśloności świata rzeczywistego. Budowa *fuzzy OLAP*, będzie przełomem w technologiach hurtowni danych i pozwoli za pomocą rozmytych mo-

deli matematycznych opisać oraz analizować informacje dotyczące działalności organizacji, także w warunkach niepewności.

### Zakończenie

Pojawia się nowa szkoła w naukach organizacji i zarządzania – szkoła informacji i wiedzy. Właściwe przetworzenie ogromnej ilości informacji z wielu różnych dziedzin oraz zarządzanie wiedzą wydobywaną z tych informacji może być efektywne jedynie w oparciu o modelowanie matematyczne.

Istniejący od lat metodologiczny spór w teorii podejmowania decyzji: podejście analityczne oparte o modele podejmowania decyzji i koncepcja decyzji deliberacyjnych, opartych na całościowej analizie danych, zapoczątkowana od ogólnej teorii systemów Bertalanffy’ego, którego myśli kontynuowała angielska szkoła tzw. miękkiego podejmowania decyzji – stracił obecnie na ostrości. Nadchodzące lata to epoka wiedzy i przetwarzania informacji. Tylko rozwój informatyki a szczególnie zintegrowanych systemów informatycznych pozwala na stosowanie w praktyce gospodarczej zarządzania informacją i wiedzą.

Budowanie trwałej przewagi konkurencyjnej wymusza na przedsiębiorstwach zdobywanie wiedzy a to wymaga przetwarzania informacji. Współczesny system informacyjny jest systemem skomputeryzowanym, w którym zamodelowano wielowymiarowe struktury informacyjne.

Sztandarowe pozycje przedstawicieli tzw. miękkiej analizy systemowej, jak np. *Umysł ponad komputerem* H. i S. Dreyfusów, w której przekonują, że specjaliści wysokiej klasy mogą podejmować decyzje intuicyjnie – utraciły na aktualności. Coraz powszechniej pojawiają się głosy, że tylko zastosowanie komputerów pozwala na przyjęcie właściwej strategii rozwoju organizacji.

Modelowanie matematyczne jest podstawą integracji systemu informacyjnego zarządzania poprzez przyjęcie niezbędnych formalizmów. Konstrukcja modelu matematycznego dla organizacji musi spełniać warunki integracji wszystkich jego różnorodnych elementów, hierarchii, atrybutów z uwzględnieniem oddziaływań oraz powiązań z otoczeniem.

Zrozumienie i kształtowanie przyszłości organizacji XXI w., to sztuka tworzenia i weryfikacji kompleksowych, wielowymiarowych modeli matematycznych.

### Bibliografia

- Agrawal R., Gupta A., Sarawagi S., *Modeling multidimensional databases*, Proc. 13<sup>th</sup> ICDE Conference, IEEE, Birmingham 1996.
- Chaudhuri S., Dayal U., *An overview of data warehousing and OLAP technology*, ACM SIGMOND Record, 1997, Vol. 26.
- Chojnacki A., *Modelowanie matematyczne*, WAT, Warszawa 1986.
- Codd E.F., Codd S.B., *Providing OLAP to User-Analyst*, Codd & Associates (<http://www.arborsoft.com/papers/coddTOC.html>)

- Devlin B.A., Murphy P.T., *An Architecture for a Business and Information System*, IBM Systems Journal, 1988, Vol. 27, No. 1.
- Frączek J., Gorawski M., Kozielski S., *Modelowanie struktur wielowymiarowych w hurtowniach danych*, Arch. Informatyki Teoretycznej i Stosowanej PAN, 2000.
- Gorawski M., *Systemy DSS: Hurtownia Danych*, „Informatyka” nr 3/2000.
- Grzegorzewski P., *Wspomaganie decyzji w warunkach niepewności. Metody statystyczne dla nieprecyzyjnych danych*, Exit, Warszawa 2006.
- Hand D., Mannila H., Smyth P., *Eksploracja danych*, WNT, Warszawa 2005.
- Hansen M.T., *Knowledge Management's Next Generation*, HBR, Mar. Inmon W.H. (1993) *Building the Data Warehouse* QED/Wiley, 2001.
- Jarke M., Lenzerini M., Vassiliou Y., Vassiliadis P., *Hurtownie danych: podstawy organizacji i funkcjonowania*, WSiP, Warszawa 2003.
- Kimball R., *The data warehouse toolkit: practical techniques for building dimensional data warehouse*, John Wiley & Sons, Inc. 1996.
- Kimball R., Reeves L., Ross M., Thornthwaite W., *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses* John Wiley & Sons, Inc. 1998.
- Kisielnicki J. (red.), *Zarządzanie wiedzą we współczesnych organizacjach*, WSHiP, Warszawa 2003.
- Koźmiński A.K., *Jak stworzyć gospodarkę opartą na wiedzy*, [w:] *Strategia rozwoju Polski u progu XX wieku*, Komitet Prognoz „Polska 2000 Plus”, PAN, Warszawa 2001.
- Larose D.T., *Odkrywanie wiedzy z danych*, WN PWN, Warszawa 2006.
- Łachwa A., *Rozmyty świat zbiorów, liczb, relacji, faktów, reguł i decyzji*, Elit, Warszawa 2001.
- Poe V., Klauer P., Brobst S., *Tworzenie hurtowni danych*, WNT, Warszawa 2000.
- Thomas H., Datta A., *A Conceptual Model and Algebra for OLAP*, Information Systems Research, 2001, vol. 12, No. 1, 20.
- Todman Ch., *Projektowanie hurtowni danych – zarządzanie kontaktami z klientami*, WNT, Warszawa 2003.
- Zadrozny S., *Zapytania nieprecyzyjne i lingwistyczne podsumowania baz danych*, Exit, Warszawa 2006.

## On-Line Analytical Processing (OLAP) as a Information Tool for Modern Management

---

### Summary

*The article contains on question, what it is the wholesalers of data and what are the advantage from its initiating. It defines the actual information needs for management as well as it defines the part of wholesalers of data in satisfying these needs.*

*It the conception of multidimensional structures of data, in the article formalize multidimensional model of data and basic structure wholesalers data are introduced, built according to authoress' own conception. It is the OLAP cube. Publication contains also the proposal of innovation and it forecasts the rise of new generation of wholesalers data for solving modern management problems.*