

Historia baz danych, hurtowni danych oraz modeli danych (w zarysie)

-v.2.2 Copyright by Zygmunt Ryznar

1959	Powstaje konsorcjum CODASYL ("Conference on Data Systems Languages"), którego zadaniem był nadzór i stymulowanie rozwoju standardów języków opisu i manipulacji danych. Organizacja zajmowała się początkowo głównie językiem COBOL a potem w ramach DBTG (Data Base Task Group) również sieciowymi i hierarchicznymi bazami danych ze strukturami łańcuchowymi (języki DDL - Data Description Language, DML - Data Manipulation Language, schematy i podschematy). Pojawienie się relacyjnych baz danych i ich popularność doprowadziły do zaniku działalności tej organizacji. Wg standardów "codasylofskich" opracowano co najmniej kilka szeroko stosowanych systemów zarządzania bazami danych np. IDMS.	T-MD
Lata 60-te	Sieciowe modele baz danych (wg zaleceń CODASYL-DBTG) i hierarchiczne (gł. IMS firmy IBM)	
1963	Ch. Bachman w GE opracowuje IDS (Integrated Data Store), który legł u podstaw standardu CODASYL-DBTG	S-COD
1969	Cincom Systems oferuje bazę TOTAL	S-COD
1969	IBM wprowadza do sprzedaży IMS	S-HIE
1969-1973	Powstaje IDMS (B.F. Goodrich, Culliname) na bazie języka IDS. Po przejściu Culliname w latach 90-tych CA implementuje w systemie SQL i XML, przyczyniając się do utrzymania systemu na rynku po dzień dzisiejszy.	S-COD
1971	SAG (Software AG) wprowadza bazę ADABAS (z listami inwersyjnymi)	S-INV
1970-1972	E.F. Codd opracowuje model relacyjnej bazy danych	T
1972 lata 80-te	Jay Wurtz i Rick Karrash wówczas studenci Sloan Management School, a potem twórcy pakietu Express (pierwszego produktu pracującego tak naprawdę na technologii olapowej - choć ten termin jeszcze wówczas nie istniał), zapoczątkowali podstawy teorii wielowymiarowych baz danych. Pakiet software'owy bazy Express został potem stworzony w MDS (Management Decision Systems) na IBM mainframe jako Mainframe Express, a następnie w latach 80-tych w różnych wersjach (np. pcExpress) upowszechniany przez IRI (Information Resources, Inc.). Przejęty przez ORACLE w 1995 r.	S-HD
1974	Powstaje 1sza niekomercyjna relacyjna baza danych Ingres stworzona w U.C. Berkeley - z linii tej wywodzą się Sybase, MS SQL server. Uniwersytet Berkeley był jednym z głównych ośrodków badawczych nad relacyjnymi bazami danych (Ingres i język Quel) i to jego naukowcy utworzyli firmę Relational Technology, promującą Ingres. Quel się nie przyjął (choć był bardziej strukturalny), a Sequel (SQL).	S-REL
1976	P. Chen proponuje model Entity-Relationship (ER) do modelowania baz danych	T-MD
1974-1978	IBM od 1974r opracowuje system R wprowadzając do niego potem język SEQUEL. Z linii tej wywodzą się DB2, Oracle, HP-Allbase, Tandem-Non-Stop SQL	S-REL
1978	D. D. Chamberlin and R.F. Boyce w IBM opracowali język SEQUEL (Structure English Query Language) potem przemianowany na SQL	T
1979	Powstaje komercyjna relacyjna baza ORACLE z SQL. Teoria Codd'a czekała prawie 10 lat na komercyjną realizację. W 1979 roku Larry Ellison - założyciel firmy Relational Software (przemianowanej 6 lat później na ORACLE) stworzył wraz z Bobem Minerem i Edem Oatesem pierwszą komercyjną relacyjną bazę ORACLE.	S-REL
1979-1980	W Polsce opracowano (W. Staniszkis i A. Dutkowski) RODAN. Podążając zgodnie z rozwojem światowej technologii w Rodanie utworzono potem moduł komunikacji w języku SQL do "mapowania" modelu relacyjnego.	S-COD
1981	INFORMIX Sippl i King tworzą INFORMIX (INFORMATION on unIX) w firmie Relational Database Systems (RDS), w 1985 r wprowadzają Informix-SQL. Potem RDS zmienia nazwę na Informix Software. W 2001 roku firma zostaje kupiona przez IBM.	S-REL
1981	Komercyjna wersja szbd Ingres	S-REL
1983-1985	W IBM powstaje relacyjna baza danych DB2	S-REL
1984	Teradata tworzy tzw. równoległe (parallel) bazy danych dla data-warehouses i data marts. Potem już w ramach NCR w 2000 roku oferowany był pakiet pod nazwą Teradata Warehouse Suite, składający się z takich narzędzi wspomagających eksploatację hurtowni jak: TeraMiner, Database Query Manager, Teradata Performance Pak, Teradata Manager, Parallel Data Pump (Tpump), TeraMart, Meta Data Services itp.	S-REL HD
1984	W Polskim MSW opracowano System Zarządzania Bazą Danych JANTAR	XX
1984	SYBASE Założona w Berkeley firma produkująca relacyjne systemy bazodanowe z data warehouse (Warehouse Studio). W 2012 r wchłonięta przez SAP.	S-REL HD

Po 1985 r.	Object Oriented Data Modelling - Modele Danych zorientowane obiektowo. Po niezbyt udanych próbach stworzenia oryginalnych obiektowych baz danych model zaimplementowano potem w relacyjnych bazach	T-DM
1986	ANSI publikuje standard języka SQL, potem akceptuje go ISO	T
1993	OLAP (On Line Analytical Processing) - przetwarzanie analityczne w trybie on-line. Termin (wraz z towarzyszącymi mu 12 zasadami) utworzony 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.	T
1995	MySQL AB (Szwecja) opracowuje system zarządzania bazą danych MySQL	S-REL
po 1997 r.	użycie XML do internetowego dostępu do baz danych.	
1998	NoSQL terminu tego po raz pierwszy użył Carlo Strozzi, lecz nie oznaczał on jeszcze pełnego oderwania od relacyjności	
2009	NoSQL Johan Oskarsson ponownie wprowadza ten termin do oznaczenia w pełni nierelacyjnych rozproszonych baz danych, nie realizujących reguły ACID(atomicity, consistency, isolation, durability) i stosujących różnorodne zasady budowy baz danych: key-value stores, document databases, wide column stores, graph databases.	No Rel
2015	najpopularniejsze bazy NOSQL MongoDB, Apache Cassandra, Redis	

Legenda:

System zarządzania bazą danych	S	
Praca teoretyczna	T	
Narzędzie	N	
Typ systemu zarządzania bazą danych	COD	Wg schematu CODASYLu
Typ systemu zarządzania bazą danych	REL	Relacyjna baza - wg modelu Codda
Typ systemu zarządzania bazą danych	HIE	Hierarchiczna
Typ systemu zarządzania bazą danych	VEC	Wektorowa
Typ systemu zarządzania bazą danych	INV	Inwersyjna
Typ systemu zarządzania bazą danych	XX	Inna
Modelowanie danych	MD	
Hurtownia danych	HD	
Data Mining	DM	

WYBITNI TWÓRCY

William Oille	Od 1953 r. pracuje na jednym z pierwszych komputerów Mark 1 w Manchester Univ. W 1957 r. uzyskuje tytuł doktora z astrofizyki. Od początku lat 70-tych aktywny w organizacji CODASYL jako przewodniczący komitetu d/s systemów oraz wiodący autor raportów "Generalized Database Management Systems". Specjalizował się w zastosowaniach baz danych i metodologii systemów informacyjnych. Reprezentował British Computer Society w IFIP TC8 i był aktywny w ISO (w BSI standards committee) w zakresie standardów dbms. Po przejściu na emeryturę interesował się historią rozwoju informatyki. Wiodący współautor książki "Information Systems Methodologies" (1988, 1991 401 str. Addison-Wesley Publ.Co), w której na tle ogólnej metodologii projektowania systemów informacyjnych przedstawiono zagadnienie modelowania danych w biznesie w aspekcie wykorzystywania systemów baz danych.
Edgar.F.Codd	Zasługą E.F.Codda było stworzenie podstaw relacyjnych baz danych, które stały się siłą napędową obecnego dużego biznesu bazadanowego. Ale nie przyszło mu łatwo. Jego teorie na przełomie lat 60/70 tych za bardzo wyprzedzały ówczesną technologię, kiedy królowały sieciowe (np. typu DBTG Codasyli) i hierarchiczne bazy danych (np. IMS firmy IBM). A przecie ż nie proponował rzeczy bardzo złóż onych, lecz proste tablice składające się z wierszy i kolumn, które obudował pewną notacją oraz regułami tzw.normalizacji przeciwdziałającej dublowaniu danych. . Zasługą jego jest to, że uznał iż w obsłudze zapytań lepiej stosować relacje oparte na wartości danych, niż popularne wówczas "łańcuchowanie" danych poprzez pointery wbudowane w rekordy. Po prostu nie uwierzono w tak proste rozwiązanie. Stworzenie pojemnych pamięci operacyjnych i dyskowych umożliwiło efektywną realizację tej technologii. Studiował matematykę i chemię na uniwersytecie w Oksfordzie. W czasie II wojny światowej służył jako pilot w siłach powietrznych Wielkiej Brytanii. Po wojnie przeniósł się do USA. Pracował na początku jako programista w IBM. Na uniwersytecie stanu Michigan w Ann Arbor jako stypendysta IBMu uzyskał tytuł doktora "computer-science", a następnie zamieszkał w Kalifornii, gdzie ponownie podjął pracę dla IBM. W 1970 wydał pracę "A Relational Model of Data for Large Shared Data Banks", w której przedstawił relacyjny model zarządzania bazami danych. Pracodawca jego - IBM - nie przyjął z entuzjazmem jego pracy i nie kwapił się z jej praktycznym

	<p>wykorzystaniem, kierując się prawdopodobnie interesem dotychczasowej sztandarowej bazy danych IMS. IBM nawet odsunął Codd'a od prac nad Systemem R i nie wykorzystał jego relacyjnego języka Alpha, lecz opracowywał język Sequel. Codd uważał ten język za niezgodny ze swoją teorią, co w efekcie zaowocowało odejściem z IBM i założeniem (wspólnie z Chrisem Date) firmy consultingowej. W latach 90-tych z powodu pogorszenia stanu zdrowia przestaje pracować.</p> <p>Codd ma też poważne zasługi w innej dziedzinie bazodanowej, dotyczącej hurtowni danych, a więc technologii wielowymiarowych danych To on w 1993 roku ukuł termin OLAP (online analytical processing) i sformułował dwanaście reguł tej technologii. Wprowadzenie OLAPu zmodernizowało relacyjne bazy danych o schematy gwiazdy (star) i śnieżynki (snowflake), służące do tworzenia tzw. kostek wielowymiarowych.</p>
Charles Bachman	<p>W 1963 r. twórca praktyk Ch. Bachman w GE wpada na pomysł sieciowych struktur danych wprowadzając relacje "łańcuchowe" (chain) - najpierw w języku firmowym GECOM a potem poprzez wzbogacenie stosunkowo sztywnej składni języka COBOL zwrotami RETRIEVAL VIA CALC CHAIN, PLACE NEAR nazwa-danej CHAIN, STORE, RETRRIEVE, MODIFY (modyfikuj ze zmianą łańcucha), DELETE (usuń z łańcucha). Swoje uzupełnienie nazwał IDS(Integrated Data Store) i legło ono u podstaw standardu CODASYL-DBTG. 1szy laureat nagrody Turinga bez tytułu doktorskiego.</p>
Bill Inmon	<p>Uważany jest za twórcę koncepcji hurtowni danych. W latach 90-tych twórca firm Prism Solutions i Pine Cone Systems tworzących software do administrowania danymi w środowisku hurtowni danych. Sformułował on takie cechy hurtowni danych jak zorientowanie na podmioty (np. produkty, klienci), uwzględnienie wymiaru czasu, nieulotność i integralność danych, ukierunkowanie na wspomaganie decyzji nie zaś na obsługę operacyjną (transakcyjną). Opracował architekturę systemu hurtowni danych, składającą się z globalnej i eksploracyjnej hurtowni danych oraz datamartów. Podkreślał znaczenie metadanych dla zachowania integralności na każdym poziomie.</p>
Sean Kelly	<p>Współautor metody Hadden-Kelly budowy hurtowni danych. Proponował tworzenie specjalizowanych biznesowo hurtowni danych (Vertical - Packaged Data Warehouse). Znany i doceniany w Europie irlandzki konsultant d/s hurtowni danych. Autor książek "Data warehousing – the route to mass customisation" i "Data Warehousing in Action".</p>
Ralph Kimball	<p>Wkład w opracowanie schematu Star schema (Star Workstation w Xerox). Autor metodologii, zwanej przestrzennym modelowaniem danych (dimensional data modeling) wykorzystywanej w budowie wielowymiarowych hurtowni danych oraz modelu macierzowego ("matrix") architektury datamartów. W swoich metodologiach łączy "techniczne" informatyczne podejście z orientacją na wspomaganie decyzji biznesowych (BI - Business Intelligence). W 1986 roku Kimball założył Red Brick Systems, gdzie do 1992 roku piastował stanowisko dyrektora (obecnie firma jest własnością IBM).</p>
Earl Hadden	<p>Współautor metody Hadden-Kelly do budowy hurtowni danych. Autor szkoleń i znany konsultant d/s budowy hurtowni danych.</p>
Larry Ellison	<p>Twórca i wieloletni szef imperium bazodanowego ORACLE (nazwa Oracle pochodzi od projektu bazy danych dla CIA, który Ellison realizował na początku kariery zawodowej w firmie Ampex), a równocześnie miłośnik jachtów, odrzutowców i miliarder słynący z wystawnego stylu życia. Ekspansywny przedsiębiorca informatyczny o potencjale i bogactwie porównywalnym z Billesem Gatesem i Stevem Jobsem. Zaczynał jako programista mainframe'u w firmie Amdahl, ale później ten kreatywny i uzdolniony biznesowo człowiek przeszedł na przedsięwzięcia o dużej skali. Dzięki sprzyjającym okolicznościom, a mianowicie przeoczeniu firmy IBM skoncentrowanej do roku 80-tego tylko na mainframe'ach i nie doceniającej koncepcji relacyjnych baz danych autorstwa swego pracownika E.Codda, firmy Microsoft, Apple i Oracle mogły uruchomić własne biznesy na minikomputery i komputery osobiste. Biznes relacyjnych baz danych rozwijał się wyjątkowo szybko. Larry Ellison dbał o własne firmy i przejmował wszystko co mogło mu pomóc (np. firmy Siebel, Hyperion, Siebel Analytics, Peoplesoft z zastosowaniami baz danych) albo zaszkodzić (zwalczanie konkurencji). Pod koniec 2014 roku Larry Ellison ustąpił ze stanowiska prezesa Oracle, pozostając w firmie jako jej właściciel i CTO.</p>

Kilka z tych postaci znałem osobiście:

- *William Olle - na konferencji IFIP w Bonn-Kolonii (czerwiec 1979) wygłaszałem referat w sprawie strukturalnego projektowania i zostałem przez W.Olle zaproszony do udziału w pracy grupy roboczej WG8. Bardzo otwarty i sympatyczny.*
- *Bill Inmon - byłem słuchaczem jego 2 referatów na konferencjach międzynarodowych. W foyer zawsze otoczony gromadką uczestników zadających pytania.*
- *Sean Kelly - przystojny stosunkowo młody (wówczas) Irlandczyk - niesłychanie sympatyczny. Poznałem go u schyłku lat 90-tych na konferencji w Barcelonie i potem w Warszawie na seminarium: "Data Warehousing: The Route to Mass*

Communication". Uciałem sobie z nim kilka dłuższych dyskusji.

- *Earl Hadden - uczestniczyłem w kursie prowadzonym przez niego w Warszawie w kwietniu 1999 r "Planning successful Data Warehouses and Data Marts". Bardzo rzeczowy i kompetentny.*

DEFINICJE

<p>HURTOWNIA DANYCH (data-warehouse)</p> <p>Hurtownia danych jest to nietransakcyjna zintegrowana wewnętrznie baza danych przeznaczona do przechowywania informacji w długim horyzoncie czasowym oraz w wielowymiarowych układach analitycznych i ukierunkowana na wyszukiwanie informacji bezpośrednio przez końcowych użytkowników. Rozróżnia się scentralizowane hurtownie globalne (jedna hurtownia globalna w skali firmy gromadząca wszystkie informacje niezbędne do potrzeb zarządzania). tematyczne hurtownie globalne (gromadzące informacje w skali firmy lecz wyselekcjonowane np. dla potrzeb pionu organizacyjnego) oraz datamarty ukierunkowane na obsługę specjalizowanych zespołów problemowych lub departamentów. W globalnych hurtowniach dane gromadzone są w skali firmy nie wg aplikacji operacyjnych (typu obsługa depozytów, kredytów...) lecz wg takich podmiotów (subjects) jak klient, produkt itp. zgrupowanych wg podobieństwa oraz w rozbudowanym układzie czasowym (informacje szczegółowe 2-3 lata, zagregowane – wiele lat) i agregacjach. W zintegrowanym rozwiązaniu globalne hurtownie korzystają z centralnego repozytorium metadanych. W datamartach dane gromadzone są wg kompetencji i zainteresowań zawodowych określonych grup użytkowników.</p>
<p>DATAMARTY</p> <p>Są to podhurtownie wydzielone z globalnej hurtowni danych (zwane "dependent datamarts") albo pseudohurtownie będące wyciągami z operacyjnych baz danych (zwane "stovepipe" lub "independent datamarts") nie stowarzyszone z metadanymi lecz wykorzystywane przez narzędzia OLAPowe. Datamarts ukierunkowane są biznesowo na obsługę wydzielonych grup użytkowników. Nadają się do obsługi lokalnych potrzeb informacyjnych (wyrażanych zwykle w gotowych agregatach danych oraz wskaźnikach) i dają użytkownikom lepszy komfort pracy (poruszanie się po znanych merytorycznie zasobach informacyjnych, krótszy czas odpowiedzi itp.). Stosuje się je w celu zmniejszenia nakładów (mniejsze komputery, prostsza i szybsza budowa) lub minimalizacji ruchu danych po sieci (wtedy noszą charakter zdecentralizowany czyli ulokowane są na serwerach departamentowych). W stosunku do hurtowni danych datamarts powinny być więc prostsze pod względem modelu danych oraz posiadać mniejszy wolumen danych, zwykle sięgający kilkunastu - czasem kilkudziesięciu GB. Dane pochodzą zwykle tylko z kilku systemów, zaś liczba użytkowników nie przekracza 100. Niezależne datamarty stosuje się niekiedy w początkowym etapie budowy hurtowni danych w celu wypróbowania na nich narzędzi OLAPowych lub też do przechowywania wyników zapytań ad-hoc albo złożonych przekroi informacyjnych (cross-subject analysis).Większość specjalistów uważa, iż stosowanie wielu niezależnych datamartów może doprowadzić do chaosu informacyjnego ("wiele prawd"), jest w sumie kosztowne i utrudnia budowę globalnych hurtowni danych.</p>
<p>DATA-MINING "inteligentna eksploracja danych" mająca na celu pozyskiwanie wiedzy ukrytej w dużej ilości danych (dosłownie "kopanie danych").</p> <p>Termin ten utożsamiany jest czasem z "data drilling". W terminologii firmy SAS "data-mining" oznacza procesy selekcji, eksploracji i modelowania, wykonane na dużej ilości danych, prowadzące do odkrycia dotychczas nieznanymi wzorców (patterns) biznesowych. Sean Kelly - autorytet w zakresie hurtowni danych, określa "data mining" jako nietrywialną ekstrakcję poprzednio nieznaną wiedzę z danych, przechowywanych w hurtowni, i oznacza to skrótem KDD (Knowledge Discovery in Databases). Wreszcie, wg Johna Mangolda, prezydenta firmy Angoss (pakiety KnowledgeSeeker i KnowledgeStudio) "data mining" można zdefiniować jako proces wykrywania korelacji i trendów w danych oraz dostarczania wiedzy, za pomocą technik statystycznych, matematycznych i rozpoznawania wzorców. Wg interpretacji Gartner Group "Data mining jest procesem odkrywania nowych korelacji, wzorców i trendów na podstawie dużych wolumenów danych przechowywanych w repozytoriach, wykorzystując technologie rozpoznawania wzorców, jak również techniki statystyczne i matematyczne". Wg firmy IBM data mining jest procesem wydobywania z dużych baz danych nieznanej do tej pory informacji i następnie używanie jej do podejmowania istotnych decyzji biznesowych.</p>
<p>OLAP (On Line Analytical Processing) - przetwarzanie analityczne w trybie on-line ,w skrócie: przetwarzanie analityczne Termin (wraz z towarzyszącymi mu 12 zasadami) utworzony w 1993 r. przez twórcę teorii relacyjnych baz danych E.F.Codda. Oto reguły OLAPu opublikowane w "białym raporcie" (white paper) firmowanym przez Comshare autorstwa E.F.Codd, S.B.Codd, C.T.Saleley :</p> <p>Wielowymiarowe spojrzenie (view) konceptualne Transparentność (niezależność od platform) Dostępność pozwalająca na utworzenie spójnego spojrzenia użytkownika (OLAP powinien dokonywać wewnętrznie konwersji tego 'view' na swoje wewnętrzne struktury danych) Wydajność raportowania niezależna od wzrostu liczby wymiarów Serwer OLAPowy działający w architekturze klient-serwer z dostępem przyjaznym dla klientów Równość wymiarów: struktury danych i funkcje mogą być przydzielane dowolnym wymiarom Tylko jeden optymalny fizyczny schemat danych dla każdej macierzy danych Narzędzie OLAPowe dostępne równocześnie dla wielu użytkowników wykorzystujących te same dane lub te same analityczne modele. Nieograniczone operacje międzywymiarowe (unrestricted cross-dimensional operations): możliwość przyporządkowania obliczeń do dowolnych wymiarów z możliwością konsolidacji przy przechodzeniu pomiędzy poziomami agregacji,</p>

dozwolone relacje pomiędzy dowolnymi pozycjami danych (data cells) Intuicyjne "drażenie" danych (np.drill down) - bez przechodzenia przez aparat pośredniczący (menu lub inne łącze) Elastyczne raportowanie - wg wszystkich wymiarów włączając wszystkie możliwe podzbiory. Nieograniczona liczba wymiarów i poziomów agregacji, a przynajmniej 15 wymiarów (najlepiej 20) stowarzyszonych ze wspólnym modelem analitycznym. Każdy wymiar powinien mieć nieograniczoną liczbę agregacji definiowanych przez użytkowników w ramach danej ścieżki konsolidacyjnej. 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. Z punktu widzenia końcowego użytkownika OLAP oznacza przede wszystkim wielowymiarową analizę danych zainicjowaną 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 Technika OLAP działać może na relacyjnych bazach danych (ROLAP- Relational OLAP) lub specjalnie zaprojektowanych bazach wielowymiarowych (MOLAP-Multidimensional OLAP). MOLAP osiągnąć jest np. poprzez wektorowe (macierzowe) struktury danych lub oparte na geometrii fraktalnej, zaś ROLAP poprzez "nakładki" na relacyjne bazy danych (np. Star Schema, bitowe mapy indeksowe, fizyczna segmentacja danych wg wymiarów, dynamicznie tworzony wielowymiarowy model danych). Do MOLAP zaliczane są przede wszystkim Essbase (Arbor/Hyperion), Acumate (Kenan), Express(Oracle), MDDDB (SAS) i TM/1 (TM/1). Pod koniec lat 90-tych niektóre narzędzia (np. Holos, Express, OLAP Services SQL Server) zostały rozbudowane do postaci hybrydowej (zasługując na nazwę HOLAP - Hybrid OLAP), obsługującej zarówno bazy relacyjne jak i kostki wielowymiarowe. Tak więc w hurtowniach mogą być stosowane różnorodne struktury danych: klasyczne znormalizowane, znacznie zmodyfikowane (zdenormalizowany schemat gwiazdzisty lub wielogwiazdzisty, "pośredni" schemat śnieżynki-"snowflake schema" polegający na dekompozycji wymiarów) lub też zupełnie oryginalne w postaci tzw. kostek (sześciiany "multicube" w pakietach Express,TM1 i "hypercube" w Essbase). W kostce Expressowej przechowywane są wartości miary (measures), np. sprzedaży, w układzie hierarchii wymiarów. W bazie danych definiuje się wiele kostek, które korzystają z wspólnych metadanych definiowanych na poziomie całej bazy. Wymiana (operacje eksportu i importu) danych pomiędzy bazami odbywa się za pośrednictwem specjalnych plików (EIF-Express Intechange Format), zawierających nie tylko wartości zmiennych lecz również stowarzyszone z nimi wymiary (w przypadku eksportu obliczanej zmiennej przekazywane są tylko wymiary i wzór zmiennej). Innym typem wielowymiarowości są indywidualnie wymiarowane kostki (infocubes) hurtowni biznesowej BW (Business Warehouse) firmy SAP. Firma SAS oferuje wielowymiarową bazę danych MDDDB, w której kostka wielowymiarowa, obsługiwana przez mechanizm zwany "NWAY crossing", może posiadać wiele zmiennych (miar). Wiele zmiennych (do 127) posiadać może kostka w OLAP Service w ramach SQL Server firmy Microsoft. W przypadku relacyjnych baz danych wysiłek główny wydaje się być ukierunkowany na eliminowanie takiej wady jaką jest tworzenie iloczynu kartezjańskiego tablic, np. w Oracle8 osiąga się to za pomocą dynamicznie łączonych indeksów bitowych.

SYSTEM ZARZĄDZANIA BAZĄ DANYCH (SZBD)

Oprogramowanie usługowe, znajdujące się poza oprogramowaniem aplikacyjnym, odpowiedzialne za utrzymywanie relacji pomiędzy plikami, ich bezpieczeństwo w warunkach wielodostępu i awarii, wyposażone w język zapytań, generator raportów i inne narzędzia. Istnieje kilka podstawowych typów struktur utrzymywanych przez SZBD: relacyjne (oparte na standardzie SQL i uznawane za najbardziej otwarte), hierarchiczne (np. IMS firmy IBM) i binarno-sieciowe CODASYLowskie.

NOSQL

Nierelacyjne rozproszone bazy danych, nie realizujące reguły ACID(atomicity, consistency, isolation, durability) i stosujące różnorodne zasady budowy baz danych: key-value stores, document databases, wide column stores, graph databases. Używane m.i. do potrzeb Bigdata, kiedy relacyjne bazy danych nie są w stanie skalować ogromnych ilości różnorodnych i szybko namnażających się danych.

BI GDATA

W uproszczeniu: Big Data jest to Data-mining na wielkich różnorodnych bazach danych. Cechy bigdata: duża ilość danych, duża zmienność, duża różnorodność. Big data – termin odnoszący się do dużych, zmiennych i różnorodnych zbiorów danych, których przetwarzanie i analiza jest trudna ale jednocześnie wartościowa, ponieważ może prowadzić do zdobycia nowej wiedzy. Oznacza sytuację, gdy zbioru nie da się przetwarzać przy użyciu trywialnych, powszechnie dostępnych metod.