

Gabriela M. Idzikowska*

METODOLOGICZNE PROBLEMY PROJEKTOWANIA
BANKU DANYCH

1. Podstawowe pojęcia

Istnieją dwie grupy definicji banku danych. W pierwszej z nich są uwypuklone te właściwości banku danych, które są istotne z punktu widzenia technologii przetwarzania danych, podczas gdy druga grupa definicji koncentruje się na zawartości informacyjnej banku danych.

Bank danych jest rozumiany jako zespół dwóch składników¹: bazy danych i procedur umożliwiających przetwarzanie danych zawartych w bazie. Procedury programowe stanowią system zarządzania bazą danych.

Najpełniejszą definicję technologiczną bazy danych podaje J. Martin²: "Baza danych może być zdefiniowana jako zbiór powiązanych między sobą danych przechowywanych razem bez zbędnej lub szkodliwej redundancji w celu obsługiwanego jednego lub więcej zastosowań w sposób optymalny; dane są przechowywane tak, że są one niezależne od programów, które je wykorzystują; powiązane i kontrolowane podejście jest stosowane przy dodawaniu nowych danych oraz modyfikowaniu i wyszukiwaniu danych istniejących wewnątrz ba-

*Dr, adiunkt w Katedrze Informatyki UL.

¹ Wg: W. B o g u c k i, W. S t a n i s z k i s, Projektowanie banku informowania kierownictwa, Warszawa 1974, s. 9.

² J. M a r t i n, Computer Data-Base Organization, Englewood Cliffs 1975, s. 19.

zy danych". Baza danych jest więc dużym zbiorem danych i powiązań między nimi, mającym następujące własności technologiczne³:

- złożoność organizacji danych;
- minimalizacja zajętości pamięci;
- elastyczność;
- dostępność;
- integralność;
- możliwość wyszukiwania danych według różnych kryteriów;
- wielofunkcyjność.

Powstanie tak potężnego narzędzia, jakim jest bank danych, zainspirowało specjalistów w zakresie systemów zarządzania do opracowania koncepcji wykorzystania tego narzędzia w systemie informacyjnym przedsiębiorstwa. Charakterystyczne dla tej grupy jest totalne podejście do określenia zawartości informacyjnej banku danych; banki danych traktuje się jako magazyny wszystkich danych o przedsiębiorstwie⁴.

Bank danych eksploatowany w organizacji gospodarczej powinien niewątpliwie dostarczać informacji potrzebnych w zarządzaniu. Koncepcja banku danych jako magazynu wszystkich danych o przedsiębiorstwie nie jest jednakże możliwa do zrealizowania w obecnej sytuacji w dziedzinie systemów zarządzania bazą danych. J. Martin⁵ uważa, że zamiast wprowadzać pojęcie totalnego banku danych, należałoby raczej mówić o "totalnej zasadzie organizowania", która określi strategię budowy i rozwoju banku danych. Z rozważań wynika, iż bank danych na obecnym etapie rozwoju technologii przetwarzania danych oraz teorii zarządzania może zawierać dane dotyczące wybranej dziedziny działalności lub przedmiotu gospodarowania na jednym lub kilku szczeblach zarządzania w przedsiębiorstwie. Przyjęcie totalnej zasady organizowania umożliwi w przy-

³Data Base Seminar - CONTROL DATA INSTITUT, Frankfurt am Main 1975, s. 1-33.

⁴Por. m. in. Osnovy sovremennoy sistematiki (tłumaczenie z angielskiego), Moskwa 1975, s. 76; R. G. M u r d i c k, J. E. R o s e, Information Systems for Modern Management, Englewood Cliffs 1971, s. 224; Z. G a c k o w s k i, Projektowanie systemów informacyjnych zarządzania, Warszawa 1974, s. 121; A. M. M c D o n o u g h, Systemy scentralizowane - planowanie i kontrola, Warszawa 1973; D. S. A p p l e t o n, What Data Base Isn't, "Datamation" January 1977.

⁵J. M a r t i n, Principles of Data-Base Management, Englewood Cliffs 1976, s. 232.

szłości połączenie zaprojektowanych i eksploatowanych oddzielnie banków danych.

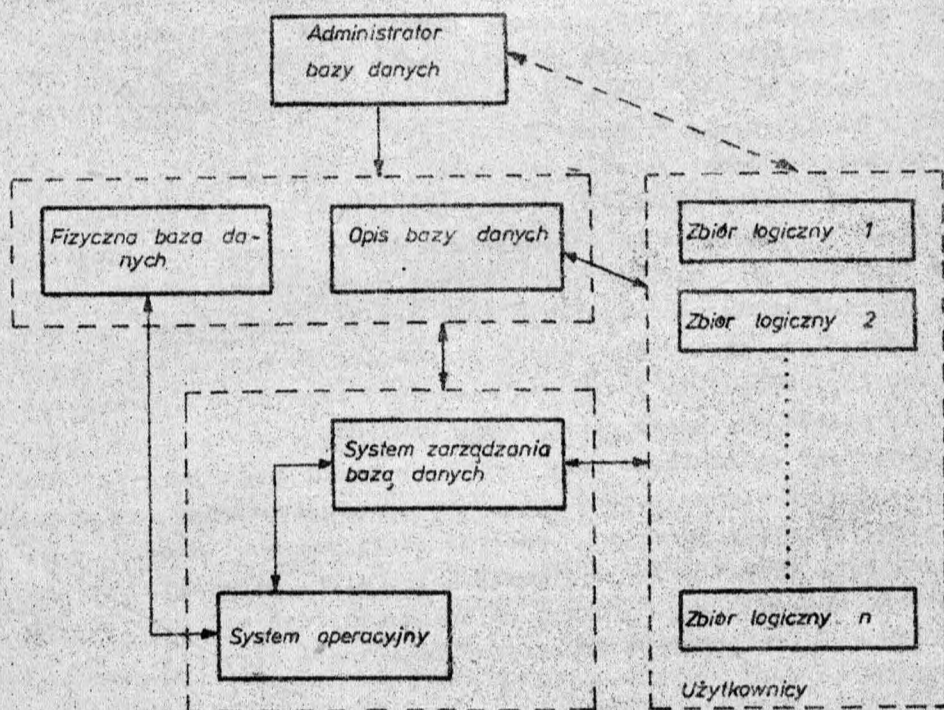
2. Baza danych

jako element systemu automatycznego przetwarzania danych

W systemie automatycznego przetwarzania danych opartym o bank danych można wyróżnić cztery elementy:

- system zarządzania bazą danych;
- administrator bazy danych;
- użytkowników banku danych;
- bazę danych.

Powiązania między tymi elementami przedstawia rys. 1.



Rys. 1. Struktura systemu informatycznego opartego na banku danych

Najważniejszą funkcją systemu zarządzania bazą danych jest obsługiwanie każdego dostępu do bazy danych. W realizacji tej funkcji współpracuje on z systemem operacyjnym komputera.

Administrator bazy danych jest strażnikiem danych zawartych w bazie: funkcję tę może spełniać jedna osoba lub grupa osób. W zakres ich obowiązków wchodzi trzy podstawowe problemy:

- organizacja bazy danych;
- bieżąca kontrola eksploatacji bazy danych;
- rozwój bazy danych.

Każdy użytkownik jest reprezentowany przez zbiór logiczny, który jest rozumiany jako podzbiór struktury logicznej bazy danych wykorzystywany przez tego użytkownika.

Jądro systemu stanowi baza danych, łącząca ze sobą wszystkie pozostałe elementy.

Elementarnym pojęciem w ramach obszaru bazy danych jest *d a n a*. Stanowi ona odzwierciedlenie atrybutu obiektu rzeczywistości. Daną⁶ określają dwie cechy: nazwa i wartość⁷. Nazwa jest cechą umożliwiającą identyfikację danej; jest ona niezależna od wartości i stała w procesie przetwarzania informacji. Wartość jest zmienną cechą danej; można ją określić jako elementarny, niepodzielny semantycznie fragment informacji.

Niezmienna cecha określa typ danej, tj. własność (atrybut) obiektu lub zbioru obiektów, rodzaj grupy danych, typ rekordu, a więc jest ona właściwa do stosowania w sferze logicznych konstrukcji bazy danych. Wartość jest zmienną cechą danej, wobec czego mogą istnieć różne reprezentacje tego samego zbioru nazw danych. Jeśli określony zbiór nazw danych elementarnych nazwie się typem rekordu logicznego, to różne reprezentacje tego zbioru nazw określa się mianem wystąpień danego typu rekordu.

Dwoistość składników bazy danych wymusza podwójne na nią spojrzenie: od strony zbioru nazw (spojrzenie logiczne) i od strony reprezentacji tego zbioru (spojrzenie fizyczne). Między elementami bazy danych występują powiązania, które określają strukturę bazy danych. Analogicznie do dwóch spojrzeń na bazę danych istnieją również jej dwie struktury: logiczna i fizyczna.

⁶ Innymi określeniami danej są: dana elementarna (ang. data element), pozycja danych (ang. data item), pole (ang. field).

⁷ W. M. T u r s k i, Struktury danych, Warszawa 1971, s.17.

Struktura logiczna lub model bazy danych odnoszą się do nazw danych i powiązań między nimi. Związki te wynikają z roli danych w systemie informatycznym i w zależności od stopnia ich złożoności wyróżnia się następujące typy modeli bazy danych: liniowy, hierarchiczny (drzewiasty), sieciowy i relacyjny.

W strukturze liniowej brakuje jakichkolwiek związków między danymi poza zależnością sekwencyjną (ang. flat files).

Struktura hierarchiczna, zwana też drzewiastą (tree), charakteryzuje się wielopoziomowością. Istnieje tu zależność "ojciec - syn", tzn. element nadrzędny może mieć wiele elementów podrzędnych, zaś element podrzędny podlega tylko jednemu elementowi nadrzédnemu.

Struktury sieciowe (networks) tym się różnią od struktur hierarchicznych, że nie obowiązują w nich zasada "ojciec - syn", a więc każdy element podrzędny może mieć wiele elementów nadrzędnych..

Baza danych zbudowana na podstawie relacji nazywa się relacyjną bazą danych. Pojęcie relacji jest zaczerpnięte z matematyki, ma wobec tego swoją matematyczną definicję: "Dane są zbiory $D_1, D_2, \dots, D_n$ (niekoniecznie różne). R jest relacją opisaną na tych zbiorach, jeśli istnieje zbiór n -elementowych ciągów danych (uporządkowanych n -tek, ang. n -tuples) $d_1, d_2, \dots, d_n$ takich, że d_1 należy do D_1 , d_2 należy do D_2 , ..., d_n należy do D_n . $D_1, D_2, \dots, D_n$ są nazywane domenami lub dziedzinami (ang. domains) relacji R . Wartość n zwana jest stopniem relacji R ⁸. W notacji matematycznej można relację zapisać w postaci wyrażenia

$$R (D_1 \#, D_2 \#, \dots, D_n)$$

gdzie $D_1 \#, D_2 \#$ są kluczowymi domenami relacji.

Struktura fizyczna bazy danych, zwana również strukturą pamięci, jest związana z rozmieszczeniem danych i powiązań między nimi na nośnikach pamięciowych w taki sposób, aby dane te mogły być niezawodnie odnalezione w późniejszym czasie. Podstawową stru-

⁸ C. J. Date - An Introduction to Database Systems, Addison-Wesley Publishing Company, 1975, s. 42.

którą, na której bazują fizyczne organizacje danych, jest lista (lub łańcuch). Z każdą listą jest związany atrybut. Lista zawiera wskaźniki⁹ rekordów, w których wartość tego atrybutu jest jednakowa. Wskaźniki mogą się mieścić w rekordzie lub poza nim. Dzięki różnym sposobom połączeń elementów struktury fizycznej powstają różne formy organizacji danych.

3. System zarządzania bazą danych

System zarządzania bazą danych (data base management system) jest drugą częścią składową banku danych. Jest on tym elementem systemu informatycznego, który udostępnia bazę danych zgodnie z ustalonymi wcześniej zasadami. Stopień rozwoju systemu zarządzania bazą danych ma znaczny wpływ na jakość systemu informatycznego, w którym on działa.

Dla komunikacji ze swoim otoczeniem system zarządzania bazą danych jest wyposażony w następujące mechanizmy pośredniczące (interfaces):

- język opisu danych fizycznych;
- język globalnego opisu danych;
- język opisu danych użytkownika;
- język manipulacji danymi.

Język opisu danych fizycznych służy do określenia fizycznego rozmieszczenia i struktury danych na nośnikach pamięciowych, dotyczy więc struktury fizycznej bazy danych.

Logiczne globalne spojrzenie administratora na bazę danych "materializuje się" do postaci użytkowej, zwanej globalnym logicznym opisem bazy danych lub schematem, za pośrednictwem języka globalnego opisu danych (języka opisu danych schematu). Schemat, utworzony za pomocą klauzul języka globalnego opisu danych, jest wykorzystywany przez system zarządzania bazą danych do wydobycia rekordów logicznych z rekordów fizycznych i do wydobycia rekordów zadaných przez programy zastosowań z rekordów logicznych.

Spojrzenie użytkownika na bazę danych może być przedstawione w postaci tzw. subschematu, utworzonego za pomocą języka progra-

⁹ Wskaźnik jest rzeczywistym, względnym lub symbolicznym adresem kolejnego rekordu powiązanego w jakiś sposób z rekordem danym.

mowania (deklaracja danych) lub za pomocą języka opisu danych subschematu dostarczonego przez system zarządzania bazą danych.

Język manipulacji danymi jest mechanizmem pośredniczącym między programem zastosowań a systemem zarządzania bazą danych. Stosunek tego języka do języka opisu danych jest taki, jak stosunek działu procedur do działu danych w programie napisanym w języku COBOL. Język manipulacji danymi stanowi zestaw makroinstrukcji lub instrukcji wywołania podprogramu (CALL) dla programów zastosowań.

Oprócz mechanizmów pośredniczących system zarządzania bazą danych musi być wyposażony w inne jeszcze mechanizmy (programowe), zwane funkcjami pomocniczymi, które materializują się w następujących programach¹⁰:

- ładowania bazy danych;
- kontroli poprawności elementów logicznej struktury;
- restrukturalizacji;
- odtwarzania zawartości bazy danych;
- analizy statystycznej.

Wszystkie te funkcje pomocnicze ułatwiają bieżącą eksploatację bazy danych, a także umożliwiają jej rozwój.

4. Projektowania informacyjne banku danych

4.1. Modelowanie bazy danych

Projektowanie banku danych powinno się rozpoczynać od analizy informacyjnej organizacji, aby następnie zbudowany model bazy danych odzwierciedlić w schemacie¹¹.

W modelu bazy danych powinny być zawarte te dane, które są najważniejsze z punktu widzenia celów użytkownika. Analiza organizacji umożliwiająca określenie zawartości informacyjnej bazy danych może mieć więc dwójaki charakter: może być zorientowana na procesy decyzyjne lub bezpośrednio na procesy produkcyjne.

¹⁰ Por. C. J. D a t e, op.cit. s. 16; W. B o g u c k i, W. S t a n i a z k i a, Podstawowe elementy architektury systemów zarządzania bazą danych, "Informatyka" 1974, nr 10; Raport grupy roboczej do spraw baz danych CODASYL, Warszawa 1977, s. 31-32.

¹¹ Por. D. C. T e i c h r i t z i e, F. H. L e c h o v e k y, Designing the Data Base, "Datamation" 1978, vol. 24, No 8.

Analiza zorientowana na procesy decyzyjne powinna ukazać rzeczywiste potrzeby informacyjne w organizacji, ustalając miejsce w systemie informacyjnym, w którym informacja jest skoncentrowana, i decyzje, do których podjęcia jest ona potrzebna. Przykładem takiej metody analitycznej jest metoda *m o d e l u b a d a w c z e g o* ("strumienia głównego") McDonougha¹². Model badawczy obejmuje cały zakres organizacji, poczynając od zdefiniowania celów i zadań - poprzez ustalenie zbioru kluczowych decyzji i informacji niezbędnych do ich podjęcia - aż do szczegółów tworzących bank danych, który służy przygotowaniu informacji dla kierownictwa. Można określić tę metodę jako dekompozycję podejmowanych decyzji. Nazwę tę usprawiedliwia hierarchiczny układ elementów modelu badawczego, którego najwyższy poziom stanowią cele organizacji, zaś najniższy - dane elementarne. Tak więc cele organizacji dekomponuje się na opisy obowiązków i odpowiedzialności, następnie na zbiory kluczowych decyzji, raporty, moduły systemu przetwarzania danych i dane źródłowe. Najniższy poziom tego modelu stanowi podstawę do identyfikacji elementów banku danych.

Przy analizie zorientowanej na procesy produkcyjne wychodzi się z założenia, że w przedsiębiorstwie istnieje jedna i tylko jedna struktura pracy (produkcyjna)¹³, która jest poziomą strukturą sieciową i obejmuje wszystkie czynności związane z tworzeniem produktu gotowego. Oprócz wspomnianej struktury przedsiębiorstwo ma również strukturę organizacyjną. Zapotrzebowania na informacje tworzą obie te struktury, jednakże dane konieczne do zaspokojenia tych zapotrzebowań istnieją tylko w strukturze pracy. Metody analizy *p r o d u k c y j n e j* polegają na dekompozycji systemu produkcyjnego i wyodrębnieniu obsługującego go systemu informacyjnego¹⁴. Dekompozycja systemu produkcyjnego może być realizowana w różnych przekrojach, w zależności od przyjętego kryterium analizy. Podobnie, tzn. wieloprzekrojowo, powinien być ujęty system informacyjny, jednakże proces ten jest informacyjny, a nie materialno-produkcyjny. Dowolny proces informacyjny składa się z czterech faz:

- przygotowania technicznego obsługiwanych procesów produkcyjnych;

¹² A. M. M c D o n o u g h, op. cit., s. 16-25 i 206-212.

¹³ D. S. A p p l e t o n, op. cit.

¹⁴ S. C h a j t m a n, Metodyczne podstawy identyfikacji i

- planowania obsługiwanych procesów;
- ewidencji stanów obsługiwanych procesów;
- regulacji obsługiwanego procesu.

W przypadku dekompozycji procesu produkcyjnego według struktury komórek produkcyjnych i funkcjonalno-informacyjnych integracja informacji o komórkach najniższych (stanowiskach roboczych) następuje w ramach komórek wyższego stopnia itd. Tworzy się w ten sposób tzw. stożek informacyjno-hierarchiczny.

Na takiej dekompozycji opiera się metoda projektowania blokowego Buschardta¹⁵. Projektowanie to rozpoczyna się od wstępnego rozpoznania organizacji w układzie funkcjonalno-strukturalnym. Polega ono na:

- wyodrębnieniu podsystemów wykonawczych;
- ustaleniu lokalnych celów podsystemów wykonawczych; wyspecyfikowaniu czynności wykonywanych w ramach podsystemów oraz określeniu towarzyszących im materiałów i nośników informacji;
- określeniu jednostek odpowiedzialnych za realizację poszczególnych działań.

Po zakończeniu wymienionych prac tworzy się trójpoziomowy model przedsiębiorstwa. Na najniższym poziomie są usytuowane podsystemy wykonawcze zwane funkcjami szczegółowymi, które mają określone wewnętrzne schematy przetwarzania oraz powiązania informacyjne i zasileniowe z zespołami funkcji znajdującymi się na wyższym i najwyższym poziomie.

Oba opisane podejścia analityczne powinny być w zasadzie stosowane łącznie. Analiza zorientowana na proces produkcyjny pozwala na ustalenie podstawowego zestawu danych o przedsiębiorstwie i powiązań między nimi, zaś analiza decyzji, odzwierciedlająca niejako żądania użytkowników, umożliwia zweryfikowanie utworzonego modelu bazy danych.

projektowania systemów informacyjnych, cz. 1-3, "Informatyka" 1977, nr 10, 11, 12.

¹⁵ E. Kolbuz, E. Kram, Wdrażanie systemów informatycznych w przedsiębiorstwie przemysłowym, Warszawa 1977, s. 37-39. Patrz również P. Siemkiewicz, Warunki informatyzacji systemu kierowania, "Wojskowy Przegląd Organizacji i Informatyki" 1976, nr 4.

4.2. Metody selekcjonowania danych i powiązań

Selekcję stosuje się w odniesieniu do wstępnego zestawu danych i powiązań ustalanych w wyniku analizy istniejącego systemu przetwarzania danych lub przez określenie żądań użytkowników na podstawie wywiadów.

Przy selekcjonowaniu danych i powiązań z punktu widzenia ich nadmiarowości można stosować metody macierzowe i relacyjne. Wybór metody selekcji zależy od postaci wykazu obiektów, atrybutów i powiązań uzyskanego w trakcie analizy systemu informacyjnego¹⁶. Oznacza to przede wszystkim, że metody relacyjne mogą być stosowane jedynie wtedy, kiedy system informacyjny został zidentyfikowany jako zbiór relacji¹⁷. W pozostałych przypadkach do celów selekcji wykorzystuje się metody macierzowe.

Przy selekcji metodami macierzowymi wychodzi się z założenia, że system można przedstawić w postaci równania¹⁸:

$$Y = F(X) \quad (1)$$

gdzie:

Y - wyjście systemu;

X - wejście systemu;

F - funkcja transformująca.

Jeśli założy się przy tym, że X i Y są wektorami, zaś F stano-

¹⁶ Z drugiej strony postać wykazu obiektów, atrybutów i powiązań jest determinowana przez dostępne oprogramowanie wspomagające proces projektowania informacyjnego bazy danych. Każdy z proponowanych przez firmy komputerowe pakietów ma "wbudowaną" metodę selekcji danych, która narzuca sposób opisu systemu informacyjnego. Por. np. Data Base Design Aid, New York 1977; R. G e r r i t s e n, A Preliminary System for the Design of DBTG Data Structures, "Communications of the ACM" 1975, vol. 18, No 10; A. N o w i c k i, Analiza systemów informacyjnych, Wrocław 1977.

¹⁷ Metody relacyjne można w zasadzie stosować pod warunkiem, że pierwotna postać opisu struktury danych w systemie informacyjnym zostanie sprowadzona do postaci relacyjnej. Jest to możliwe, ponieważ wszystkie struktury danych (poza sieciami) dają się przedstawić w formie relacji.

¹⁸ S. K ę d z i e r s k i, Macierzowy model projektu systemu informatycznego. Materiały z konferencji nt. Teoretyczno-praktyczne problemy projektowania informatycznych systemów zarządzania, Szczecin-Kołobrzeg 1978.

wł. macierz transformacji (powiązań), wówczas równanie (1) przyjmie następującą postać:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} f_{11} & f_{12} & \dots & f_{1m} \\ f_{21} & f_{22} & \dots & f_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n1} & f_{n2} & \dots & f_{nm} \end{bmatrix} \times \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \quad (2)$$

Selekcjonowanie danych odbywa się na podstawie dwóch rodzajów macierzy F : zerojedynkowej i wagowej. Macierz zerojedynkowa jest macierzą powiązań, która umożliwia eliminację danych nie uczestniczących w przetwarzaniu. Wierszom takiej macierzy odpowiadają dane wejściowe, zaś kolumnom – dane wynikowe; wykorzystanie danych w procesie przetwarzania oznacza się jedynką, zaś niewykorzystanie – zerem. Utworzoną w ten sposób macierz można interpretować następująco:

- jeżeli wiersz z nazwą danej zawiera same zera, to znaczy, że dana nie uczestniczy w tworzeniu innych danych;

- jeżeli kolumna z nazwą danej zawiera same zera, to znaczy, że żadna dana nie uczestniczy w tworzeniu tej danej; daną, której odpowiadają same zera w wierszu lub kolumnie, można wyeliminować.

Budowa macierzy wagowej¹⁹ jest kolejnym krokiem przy selekcjonowaniu danych. Macierz wagowa jest to macierz transformacji, która dotyczy tylko niezerowych elementów macierzy zerojedynkowej. Kolumnom tej macierzy odpowiadają zwykle jednostki organizacyjne instytucji, zaś wiersze charakteryzują dane wejściowe. Przecięcie wiersza i kolumny (element macierzy) stanowi liczba (waga) określająca znaczenie danej w odpowiedniej jednostce organizacyjnej²⁰. Analiza macierzy wagowej realizowana przy zastosowaniu metod statystycznych pozwala na eliminację danych najmniej znaczących w procesie przetwarzania.

Zastosowanie metod relacyjnych umożliwia dokonywanie selekcji

¹⁹ A. Nowicki, op. cit., s. 77 i n.

²⁰ Wagi mogą dotyczyć takich charakterystyk, jak funkcje pełnione przez dane, czynności przetwarzania lub liczby transformacji dokonywanych na danych.

powiązań, a przez to upraszcza model bazy danych. Można do tego celu wykorzystać dwie metody²¹: metodę dekompozycji trzeciej formy normalnej Codd'a lub metodę syntezy Barnsteina.

E. F. Codd wprowadził trzy normalne formy relacji. Aby wyjaśnić postacie relacji zapisanych w tych trzech formach, należy zdefiniować niektóre pojęcia podstawowe.

Zależność funkcjonalna między atrybutami należącymi do zbiorów A i B polega na tym, że w dowolnym momencie określonej wartości $x \in A$ (x) odpowiada tylko jedna wartość $y \in B$ (x), co można zapisać wzorami:

$$f : A \rightarrow B$$

tzn., że

$$\{(x, y) \mid x \in A \wedge y \in B\}$$

Mówi się wtedy, że Y jest zależne funkcjonalnie od X .

Zależność częściowa występuje wtedy, kiedy pewien atrybut jest zależny funkcjonalnie od podzbioru danego zbioru atrybutów. Oznacza to, że jeśli

$$f_1 : A_1, \dots, A_n \rightarrow B \text{ i } f_2 : A_1, \dots, A_m \rightarrow B$$

gdzie $m < n$, to $A_1, \dots, A_m \subset A_1, \dots, A_n$ i $A_1, \dots, A_m$ określa funkcjonalnie B , czyli B jest częściowo zależne od $A_1, \dots, A_n$. Jeśli dla zależności f_1 nie istnieje zależność f_2 , która spełniałaby wymieniony warunek, wówczas mówi się, że B jest w pełni zależne od $A_1, \dots, A_n$.

"Zależność pośrednia atrybutu A_i od zbioru atrybutów X w relacji $R(A_1, \dots, A_n)$ występuje wtedy, kiedy istnieje taki zbiór atrybutów $Y \subset \{A_1, \dots, A_n\}$, że

$$X \rightarrow Y, Y \rightarrow A_i \text{ i } Y \not\rightarrow A_i$$

gdzie A_i nie jest elementem X lub Y , a $\rightarrow$ oznacza zależność wielokrotną [...]²².

²¹ Por. G. W i e d e r h o l d, Database Design, New York 1977, rozdz. 7 i 8 lub M. G n i ł k a, Projektowanie logicznej zawartości relacyjnej bazy danych. Materiały z konferencji nt. Teoretyczno-praktyczne problemy ...

²² M. G n i ł k a, op. cit.

Pierwsza forma normalna jest taką postacią relacji, która składa się z elementarnych (tzn. nie będących zbiorami lub relacjami) wartości danych.

Druga forma normalna jest tak przekształconą pierwszą formą normalną, w której każdy atrybut nie będący kluczem jest w pełni zależny od klucza relacji.

Trzecia forma normalna tym się różni od pierwszej formy normalnej, że każdy z atrybutów nie będących kluczem jest zależny bezpośrednio od klucza relacji.

Zgodnie z metodą Codda zbiór relacji optymalizuje się według wymagań trzeciej formy normalnej, stosując następujący algorytm:

- eliminacja zależności częściowych bez utraty informacji, zgodnie ze schematem: $R(A \#, B \#, C, D)$, gdzie $A \rightarrow C \iff R_1(A \#, C)$ i $R_2(A \#, B \#, D)$;
- eliminacja zależności pośrednich według schematu: $R(A, B, C, D)$, gdzie $A \rightarrow B$ i $B \rightarrow C \iff R_1(B, C)$ i $R_2(A, B, D)$.

Metoda syntezy Bernsteina²³ przewiduje następujący algorytm umożliwiający eliminację zbędnych powiązań:

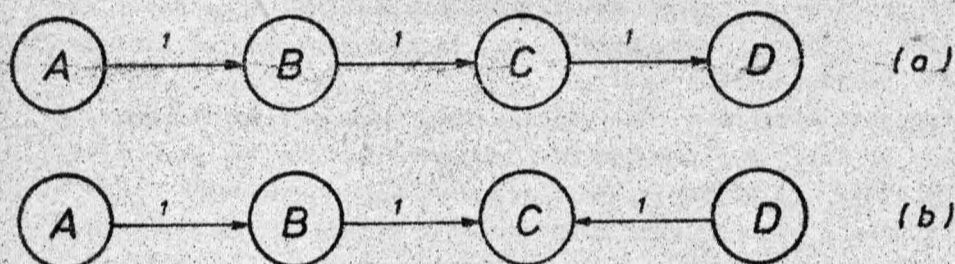
- podział zbioru zależności funkcjonalnych na grupy o identycznych lewych stronach;
- eliminacja obcych atrybutów w grupach;
- łączenie grup o identycznych kluczach;
- eliminacja zależności pośrednich;
- tworzenie relacji z wydzielonych grup zależności funkcjonalnych.

Metody macierzowe umożliwiają jedynie selekcjonowanie danych, zaś metody relacyjne eliminują powiązania częściowe i pośrednie, a więc upraszczają przedstawiony w postaci relacji wykaz obiektów, atrybutów i powiązań, który stanowi już wstępny projekt modelu bazy danych. Jest rzeczą zrozumiałą, że wstępny model bazy danych uzyskany w wyniku selekcjonowania metodami macierzowymi może zawierać redundantne powiązania. Eliminacja tych zbędnych powiązań może być dokonana metodą analizy ścieżek prostych powiązań²⁴. Ścieżkę stanowi dwu lub więcej elementowa tablica liniowa, w której między przyległymi parami elementów istnieją proste powiązania

²³ Tamże, s. 93.

²⁴ Data Base Design Aid, s. 240 i n.

(typu 1)²⁵, wszystkie skierowane w jedną stronę. Przykłady ścieżek pokazuje rys. 2.



Rys. 2. Ścieżki prostych powiązań

a) ścieżka czteroelementowa; b) przykład dwóch ścieżek prostych powiązań: ABC i DC

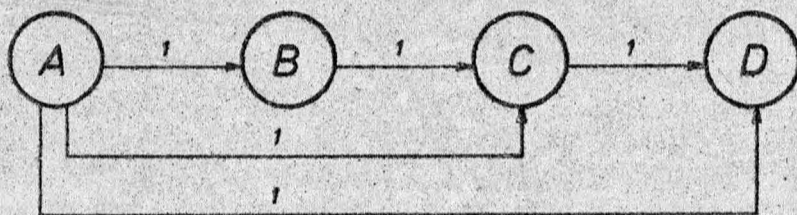
Pierwszym krokiem w omawianej metodzie jest budowa ścieżek prostych powiązań. Analiza ścieżek rozpoczyna się od klasyfikacji wyspecyfikowanych ścieżek według istniejących w ich ramach powiązań na następujące grupy:

- ścieżki z podstawowymi powiązaniem;
- ścieżki z pośrednimi (przechodnimi) powiązaniem;
- ścieżki z nowymi powiązaniem.

Powiązanie podstawowe między elementami A i B jest to takie powiązanie, które stanowi jedyną ścieżkę od A do B. Powiązania te są podstawą budowy logicznej struktury bazy danych. Powiązanie pośrednie jest rozumiane analogicznie jak pośrednia zależność w relacyjnym modelu bazy danych; takie powiązania podlegają eliminacji. Przykład powiązań pośrednich przedstawia rys. 3.

Powiązanie nowe określa się jako takie, które wynika nie z identyfikacji obiektów, atrybutów i powiązań, ale z powiązań podstawowych. To nowe powiązanie można potraktować jako takie, które nie jest w danej chwili uświadamiane przez użytkownika, ale należałoby je umieścić w bazie danych, ponieważ może ono zaspokajać przyszłe potrzeby informacyjne.

²⁵ Powiązanie typu 1 stanowi zależność "syn - ojciec" (odróżniona hierarchia). Strzałka w schemacie między A i B może być interpretowana jako: B zależy od A.



Rys. 3. Ścieżka z powiązaniem pośrednimi: AC i AD stanowią powiązania pośrednie, ponieważ można je zastąpić powiązaniem podstawowym: AB, BC i CD

Otrzymany w wyniku selekcji danych i powiązań wstępny model bazy danych stanowi podstawę do konstruowania struktury logicznej bazy danych.

Nie trudno zauważyć, że przeprowadzenie selekcji przy zastosowaniu metod relacyjnych umożliwia uzyskanie gotowego modelu bazy danych, tzn. modelu relacyjnego. W takim przypadku faza selekcjonowania danych jest równoważna z konstruowaniem struktury logicznej bazy danych.

Inaczej przedstawia się sprawa z nierelacyjnym wyrażeniem obiektów, atrybutów i powiązań. W wyniku procesu selekcjonowania uzyskuje się zestaw ścieżek danych nieredundantnych z powiązaniem podstawowymi i nowymi. Zanim powstanie na tej podstawie model bazy danych, należy ustalić klucze pierwotne i wtórne dla każdego obiektu, a następnie przeprowadzić strukturalizację zestawu ścieżek.

Zarówno klucze pierwotne, jak i wtórne określa się na podstawie przeprowadzonej analizy systemu informacyjnego. Klucz pierwotny jednoznacznie identyfikuje obiekt (rekord), przy czym kluczem może być pojedynczy atrybut (pole) lub kilka atrybutów (grupa pól). Klucz pierwotny wraz z atrybutami tworzy rekord, a identyfikując jednoznacznie obiekt, staje się jego "przedstawicielem" w tworzeniu struktury logicznej. Klucze wtórne są to te pola, które biorą udział w wyszukiwaniu informacji.

Konstrukcja struktury logicznej polega na syntezie kluczy (obiektów) i powiązań między nimi. Syntezie podlegają więc ścieżki podstawowych powiązań, w których elementami są klucze obiektów. Budowa modelu bazy danych polega na łączeniu ze sobą ścieżek z

powtarzającymi się kluczami w celu utworzenia odpowiadającej temu zestawowi ścieżek (wynikającej z niego) struktury logicznej.

Łączny model bazy danych powinien obejmować - oprócz diagramu przedstawiającego schemat struktury logicznej bazy danych - również uporządkowane informacje o wszystkich obiektach, atrybutach i powiązaniach (uzyskane z analizy systemu informacyjnego)²⁶.

"Ręczna" realizacja metody selekcjonowania danych jest niezwykle pracochłonna. Dlatego firmy komputerowe opracowały pakiety programów uprawniające projektowanie bazy danych tą metodą. Dążąc do standaryzacji dokumentacji związanej z analizą danych i analizą funkcjonalną, opracowuje się odpowiednie metody prezentacji wyników analizy (metody SADT - Structured Analysis and Design Technique oraz IA - Information Analysis)²⁷.

4.3. Weryfikacja struktury logicznej

Model bazy danych można potraktować jako swoistego rodzaju hipotezę, której prawdziwości należy dowieść: innymi słowy: konieczna jest przeprowadzenia jego weryfikacji.

Weryfikacji struktury logicznej (modelu bazy danych) dokonuje się w różnych momentach procesu projektowania bazy danych, a więc czynność ta nie stanowi oddzielnej fazy projektowania. Weryfikowanie polega na konfrontacji kolejno uzyskiwanych przybliżeń modelu bazy danych z rzeczywistością²⁸. Istnieje więc podstawa, aby stwierdzić, iż zarówno w metodzie modelowania, jak i selekcjonowanie dokonuje się weryfikacji struktury logicznej bazy danych.

Nie można jednakże założyć, że uzyskany w wyniku konstrukcji model bazy danych stanowi rozwiązanie optymalne, jeżeli nie był on konfrontowany z potrzebami użytkowników. Taka konfrontacja następuje:

²⁶ Wykaz danych o obiekcie powinien zawierać: nazwę i opis obiektu, reguły występowania, działy organizacji odpowiedzialne za występowanie obiektu oraz oczekiwaną liczebność i współczynnik wzrostu. W wykazie informacji o atrybutach występują: nazwa, opis, reguły występowania, typ danych, dopuszczalny odcinek czasu między zaistnieniem zdarzenia a aktualizacją wartości. Dla powiązań należy podać ich opis oraz liczebność.

²⁷ Por. The Analysis of User Needs, "EDP Analyzer" 1979, vol. 17, No 1.

²⁸ Takie projektowanie z jednoczesnym weryfikowaniem można o-

- w momencie zaprojektowania funkcjonalnego modelu danych;
- w momencie implementacji bazy danych;
- w czasie eksploatacji bazy danych.

Równolegle lub prawie równolegle z konstruowaniem struktury logicznej bazy danych tworzy się tzw. funkcjonalne modele danych²⁹. Są one odzwierciedleniem spojrzenia użytkowników na bazę danych, a więc wyrażają potrzeby tychże użytkowników. Projektowanie funkcjonalnych modeli danych jest oparte na wynikach analizy funkcjonalnej organizacji. Analiza ta określa wymagania dotyczące przetwarzania danych dla każdej funkcji zastosowań, co pozwala na utworzenie koncepcyjnego modelu danych związanego z tą funkcją. Konfrontacja zbioru funkcjonalnych modeli danych z modelem bazy danych umożliwia odpowiedź na pytanie: czy każdy funkcjonalny model danych stanowi podzbiór modelu bazy danych, tzn. czy model bazy danych zaspokaja potrzeby wszystkich funkcjonalnych modeli danych? Pozytywna odpowiedź na tak postawione pytanie pozwala stwierdzić, że model bazy danych jest w danym momencie kompletny.

Pojęcie kompletności modelu bazy danych jest względne, ponieważ z uwagi na długotrwałość procesu projektowania bazy danych potrzeby użytkowników mogą ulegać zmianie. Z tego powodu ponowne weryfikacji modelu bazy danych w momencie jej uruchomienia staje się koniecznością. Weryfikacja na tym etapie jest dokonywana przy wykorzystaniu testowej bazy danych. Na testowej bazie danych realizuje się działania przewidziane w systemie informatycznym opartym o bank danych. Pozwala to m. in. stwierdzić, czy struktura logiczna bazy danych jest poprawna. Niezgodności wynikiem w trakcie testowania bazy danych usuwa się przed wdrożeniem rzeczywistej bazy danych.

Ciągła weryfikacja struktury logicznej podczas rzeczywistej eksploatacji bazy danych umożliwia udoskonalenie tej struktury. Udoskonalenie struktury logicznej bazy danych może polegać na usuwaniu elementów zbędnych (martwych) lub dodawaniu nowych. Usunięcie elementów zbędnych następuje w wyniku działania funkcji pomocniczej systemu zarządzania bazą danych, która analizuje częstość

kreślić jako iteracyjne. Momenty procesu projektowania, w których dokonuje się weryfikacji, są nazywane punktami przerwań.

²⁹ Database Design. "ICL Dataskil" June 1976.

wykorzystania poszczególnych elementów struktury logicznej bazy danych. Usunięcie elementów zbędnych wymaga reorganizacji tej struktury. Dodawanie nowych elementów do modelu bazy danych jest konieczne wtedy, kiedy uległy zmianie istniejące modele danych lub kiedy pojawiły się nowe funkcjonalne modele danych. W takim przypadku musi nastąpić przebudowa modelu bazy danych. Zarówno reorganizacja, jak i przebudowa struktury logicznej bazy danych należą do obowiązków administratora bazy danych.

5. Projektowanie technologiczne banku danych

Projektowanie technologiczne wiąże opracowaną strukturę logiczną z systemem zarządzania bazą danych. Pierwszą czynnością w tej części procesu projektowania jest określenie systemu zarządzania bazą danych, które może polegać na:

- zaprojektowaniu systemu zarządzania bazą danych przy uwzględnieniu potrzeb użytkownika;
- wyborze jednego z uniwersalnych systemów zarządzania bazą danych zgodnie z ustalonymi kryteriami.

Obecnie daje się zauważyć tendencje do korzystania z uniwersalnych systemów zarządzania bazą danych (USZBD). Jest to spowodowane tym, że³⁰:

- uniwersalne SZBD oszczędzają czas wdrożenia;
- tylko w przypadku użycia uniwersalnego SZBD można wykorzystać pakiety programów zastosowań, które są na nim oparte;
- czasem trudno jest użytkownikowi sformułować precyzyjnie swoje potrzeby związane z bankiem danych;
- użytkownikowi brakuje kompetencji w zakresie problemów dotyczących banku danych;
- zastosowanie USZBD prowadzi do obniżki kosztów podstawowych związanych z wdrożeniem i eksploatacją systemu banku danych.

Odświeżenie postępowania przy określaniu systemu zarządzania bazą danych wymaga ustalenia zestawu kryteriów, które pozwolą

³⁰ T. G i l l, Data Base Software: A Sceptical Viewpoint and Some Alternatives, "Management Informatica" 1973, vol. 2, No 5.

zdecydować, czy rozważany system odpowiada wymaganiom SZBD realizującego cele użytkownika. System zarządzania bazą danych można oceniać z punktu widzenia atrybutów banku danych (integralność, kompleksowość, elastyczność, dostępność itd.). O przydatności systemu zarządzania bazą danych decydują niewątpliwie wbudowane w niego funkcje pomocnicze (kontrola poufności, zapewnienie jednocześnie dostępu do bazy danych, prowadzenie statystyk itd.). Użytkownik powinien również rozważyć czynniki wpływające na koszt zastosowania określonego systemu zarządzania bazą danych, a mianowicie³¹:

- możliwość adaptacji (portability) - w jakim zakresie system jest możliwy do eksploatacji w nowych warunkach;

- koszty wdrażania (implementability) - jaki jest rzeczywisty koszt oceny, planowania, administrowania, nauczania, programowania i monitorowania określonego SZBD, zanim zacznie on działać dostatecznie dobrze;

- możliwość utrzymywania (maintainability) - jaki jest koszt utrzymywania w stanie aktualnym badanego SZBD w porównaniu z innymi systemami;

- niezawodność oprogramowania (software reliability) - ocena kosztu i wysiłku związanego z wykrywaniem błędów w tak skomplikowanym oprogramowaniu;

- koszt działania (operational cost) - jaki jest rzeczywisty koszt użytkowania USZBD w porównaniu z innymi systemami;

- bezpieczeństwo i niezawodność danych (data security and reliability) - czy dane będą wrażliwe na zniszczenie i jaki będzie koszt odtwarzania bazy danych.

Podjęcie decyzji związanej z zastosowaniem określonego systemu zarządzania bazą danych rozpoczyna właściwe projektowanie technologiczne. O ile efektem projektowania informacyjnego był model bazy danych, o tyle w wyniku projektowania technologicznego powstaje projekt techniczno-eksploatacyjny banku danych.

W trakcie projektowania technologicznego następuje "zderzenie" modelu bazy danych oraz funkcjonalnych modeli danych z elementami systemu zarządzania bazą danych. Istnieje tyle metod w tej części projektowania banku danych, ile jest systemów zarządzania bazą danych. W każdym systemie zarządzania bazą danych działają

³¹ Tamże.

odrębne mechanizmy pośredniczące oraz funkcje pomocnicze, które wymuszają różne podejścia do projektowania. Można jednak wyróżnić fazy występujące we wszystkich metodykach technologicznego projektowania bazy danych.

Pierwszą fazą w tej części procesu projektowania jest opis struktury logicznej bazy danych, czyli utworzenie schematu. Do tego celu wykorzystuje się język globalnego opisu danych, który specyfikuje wszystkie informacje dotyczące poszczególnych elementów modelu bazy danych. W tej fazie następuje³²:

- identyfikacja danej elementarnej, grupy danych, rekordu, zbioru;
- nadanie nazwy każdemu typowi danych;
- określenie, które typy danych elementarnych i w jakiej kolejności składają się na typ grupy danych lub typ rekordu;
- określenie, które typy danych elementarnych lub kombinacji danych elementarnych są wykorzystywane jako klucze;
- nadanie nazw powiązaniom między typami rekordów;
- definiowanie sposobu reprezentacji danej elementarnej;
- określanie długości danych elementarnych;
- określenie liczby danych elementarnych, które tworzą wektor, wymiarów i wielkości macierzy lub liczby grup danych w grupie powtarzającej się;
- określenie kolejności rekordów w zbiorze lub kolejności grup rekordów w bazie danych;
- określenie zasad zabezpieczenia przed nieupoważnionym czytaniem lub modyfikacją danych.

Kolejna faza projektowania technologicznego polega na opracowaniu programów zastosowań w powiązaniu z subschematami. Faza ta stanowi zmaterializowanie drugiego nurtu projektowania informacyjnego, czyli funkcjonalnych modeli danych zbudowanych na podstawie analizy potrzeb użytkowników³³. W zasadzie pierwszą czynnością w

³² J. M a r t i n. Principles ..., s. 162.

³³ D. Appleton uważa, iż zamiast przeprowadzać wyczerpującą analizę wymagań użytkowników, aby otrzymać "zamrożoną" specyfikację potrzeb, należy szybko zbudować system na podstawie "zgrubnej" analizy i pozwolić mu się rozwijać w miarę nabywania doświadczenia i narastających umiejętności użytkownika. Patrz: Baza danych. "Europejski Program Badawczy Diebolda" Warszawa 1975, s. 110.

omawianej fazie jest zaprojektowanie programów, ponieważ one dopiero (wraz ze schematem) stanowią podstawę do opracowania subschematów. Ustalone podczas analizy potrzeby użytkowników służą do określenia funkcji użytkowych (rodzajów zastosowań)³⁴. Umożliwiają one jednocześnie określenie wejściowych i wyjściowych formatów danych. Lista danych wejściowych i wyjściowych umożliwia z kolei definicję systemu informatycznego korzystającego z banku danych; jej wynikiem jest specyfikacja procesów przetwarzania danych wejściowych na wyjściowe oraz projekt programów i przebiegów. Programy zastosowań działające na podstawie danych powstają przy użyciu klauzul i instrukcji języka manipulacji danymi, stanowiącego część systemu zarządzania bazą danych. Specyfikacja danych wejściowych dla programu określa jednocześnie postać subschematu przedstawioną w języku opisu danych subschematu odpowiednim dla wybranego systemu zarządzania bazą danych.

Ostateczny projekt subschematu powstaje w wyniku konfrontacji formy tegoż subschematu odpowiadającej wymaganiom programu zastosowań ze schematem bazy danych. Subschemat musi być bowiem tak zaprojektowany, aby:

- odpowiadał potrzebom programu;
 - nie był wrażliwy na zmiany w schemacie i strukturze fizycznej bazy danych;
 - był zgodny ze schematem i możliwy do uzyskania z niego.
- Subschemat może się różnić od schematu³⁵:
- pominięciem opisu niektórych danych;
 - zmienionymi charakterystykami danych elementarnych;
 - kolejnością danych elementarnych w grupie danych lub rekordzie oraz kolejnością rekordów w zbiorze;
 - zdefiniowaniem innych powiązań;
 - tym, że w subschemacie można opisać typy rekordów (grup danych), które są złożone z danych elementarnych zawartych w różnych typach rekordów (grup danych).

³⁴ Pojęcie "funkcja użytkowa" oznacza "proces lub sekwencję procesów obliczeniowych, w następstwie realizacji których generowany jest odpowiedni zestaw raportów pozwalający użytkownikowi na podjęcie określonych działań w przedsiębiorstwie stosownie do przyjętego zakresu użytkowego funkcji". Przykład projektowania bazy danych, Warensa 1977, s. 18.

³⁵ J. M e r t i n, Principles..., s. 163.

Trzecią fazę wyodrębnioną w procesie projektowania technologicznego jest opracowanie struktury fizycznej bazy danych. Sposób realizacji tej fazy zależy nie tylko od systemu zarządzania bazą danych, ale również od sprzętu i systemu operacyjnego. Uogólniając stosowane w tym zakresie metodyki, można wyróżnić w omawianej fazie dwie grupy czynności: opracowanie struktury fizycznej bazy danych i opis opracowanej struktury fizycznej.

Struktura fizyczna bazy danych powinna z jednej strony w sposób efektywny zaspokajać potrzeby użytkowników dotyczące dostępu do danych, a z drugiej strony - uwzględniać ograniczenia związane z wykorzystywanymi zasobami sprzętowymi. Opracowanie struktury fizycznej jest więc działaniem optymalizacyjnym. Ma ono polegać na minimalizacji pewnej złożonej funkcji³⁶:

- obciążenia pamięci pomocniczej;
- obciążenia pamięci głównej;
- czasu kanałów;
- czasu procesora.

W procesie projektowania struktury fizycznej bazy danych niezbędne są następujące informacje:

- parametry wymienionych zasobów sprzętowych łącznie z określeniem znaczenia każdego z tych zasobów;
- liczba operacji dodawania, kasowania, modyfikowania i wyszukiwania dotyczących każdego rekordu bazy danych;
- częstotliwość wykorzystania wszystkich ścieżek dostępu³⁷ w trybie partiiowym i bezpośrednim.

Projekt struktury fizycznej powinien określać poszczególne elementy tej struktury, m. in.:

- stosowane urządzenia pamięci pomocniczej;
- sposób łączenia rekordów logicznych w rekordy fizyczne;
- metody lokalizowania rekordów w bazie danych (adresowanie pośrednie i bezpośrednie);
- sposoby organizacji zbiorów fizycznych;
- metody reprezentacji powiązań między obiektami;
- zasady stronicowania w pamięci wirtualnej.

³⁶ Database Design.

³⁷ Ścieżki dostępu (ang. access paths) są określone przez powiązania stosowane między rekordami i polami, pola kluczowe, sposób uporządkowania rekordów oraz kryteria wyszukiwania informacji.

Drugą grupę czynności w fazie projektowania struktury fizycznej jest opis opracowanej struktury. Do realizacji tego zadania służy język opisu danych fizycznych, stanowiący jeden z mechanizmów pośredniczących systemu zarządzania bazą danych.

Opis struktury fizycznej bazy danych kończy proces właściwego projektowania banku danych i jednocześnie rozpoczyna jego wdrożenie. Wdrażanie banku danych (podobnie jak projektowanie) może odbywać się radykalnie lub ewolucyjnie.

Wykorzystanie metody ewolucyjnej jest możliwe jedynie wtedy, gdy w przedsiębiorstwie istnieją już systemy informatyczne; metoda ta polega na określeniu liczby i rodzajów minimalnych zmian w funkcjonujących systemach informatycznych. Taki minimalny zestaw zmian zaproponował I. A. Palmer³⁸:

- reorganizacja fizycznej struktury danych;
- zmiana logicznej struktury danych;
- zmiana dostępu do danych;
- zmiana wymagań programów indywidualnych względem danych;
- zmiana zarządzania danymi;
- wprowadzenie użytkownika systemu na bieżąco;
- kompletna zmiana wyposażenia sprzętowo-programowego systemu.

Metoda radykalna polega na zaprojektowaniu całego banku danych i wdrożeniu go od razu. Taki sposób jest droższy, a osiągnięcie pełnego wdrożenia nie zawsze pewne. Można go jednak stosować zarówno w warunkach funkcjonowania konwencjonalnych systemów informatycznych, jak i wówczas, kiedy takich systemów nie ma. Pierwszą czynnością przy wdrażaniu banku danych metodą radykalną jest przeniesienie na nośniki pamięciowe poszczególnych elementów projektu technologicznego (schematu, subschematów, programów zastosowań i opisu danych fizycznych). Wymienione elementy mogą być zapisane w pamięci w formie źródłowej lub skompilowanej³⁹.

Forma skompilowana schematu, subschematów i programów powstaje w momencie kompilacji programów: wszystkie potrzebne informacje ze schematu i subschematów są włączone do programów zastosowań.

³⁸ Za B. Wysockim, *Technologie banku danych - problemy projektowe*. Materiały z konferencji nt. Teoretyczno-praktyczne problemy ...

³⁹ G. Wiedarhold, op. cit., s. 390-391.

Forma źródłowa wiąże się z interpretacyjną metodą korzystania z banku danych: w przypadku dostępu do bazy danych program zastosowań jest traktowany jako program uniwersalny "wywołujący podprogram", czyli schemat, w którym odnajduje się potrzebne dane elementarne i ustala powiązania między nimi. "Ładowanie" (loading), czyli przenoszenie omawianych elementów na nośniki pamięciowe, odbywa się w sposób automatyczny.

Podobnie, tzn. przy użyciu specjalnie do tego przeznaczonych programów (funkcji pomocniczych systemu zarządzania bazą danych), dokonuje się ładowanie samej bazy danych (wystąpień rekordów). W początkowej fazie wdrażania wprowadza się na nośniki pamięciowe testową bazę danych w celu sprawdzenia poprawności działania systemu banku danych. Pozytywne wyniki testowania umożliwiają ładowanie rzeczywistej bazy danych i rozpoczęcie właściwej eksploatacji systemu.

Projektowanie banku danych nie kończy się w momencie uruchomienia systemu banku danych. Administrator bazy danych (reprezentujący bank danych) i administratorzy zastosowań (reprezentujący interesy użytkowników) muszą zapewniać rozwój banku danych w czasie jego eksploatacji.

Gabriela M. Idzikowska

METHODOLOGICAL PROBLEMS IN DESIGNING DATA BANK

The first part of the article deals with main concepts connected with the data bank, and more precisely with its two elements i.e. data base and data base management system.

In the second and main part of the article, the author presents her comments on propositions concerning designing of the data bank as described in the literature of the subject, and presents her own approach to the problem. The entire designing process is divided into two stages: information designing and technological designing.

Information designing aims at construction of the data base model, which can be done according to two methods described in the article: modelling and selecting.

Technological designing links the elaborated model of data base with the data base management system. There are as many methodological approaches in this part of designing the data bank as many data base management systems exist. However, some stages may

be distinguished here, which appear in all these approaches. The article specifies and describes stages in designing the technological data bank.

A separate paragraph has been devoted to the problem of verification of the data base model performed during the entire period of designing and exploiting the data bank.