

METODA TVORBY BÁZE DAT VE 4 FÁZÍCH A JEJÍ PROGRAMOVÉ ZAJIŠTĚNÍ

Ing. Jan Havlík

Ing. Martin Zralý

1. Úvod

Moderní projekční přístupy ke tvorbě informačních systémů pro řízení /a tzv. ASŘ/ usilují o racionálnější a rychlejší způsoby projektování. Přitom se vyžaduje, aby systém bylo možno vytvářet postupně s tím, že jednotlivé části jsou bezprostředně po dokončení uváděny do používání. Realizace takovýchto přístupů je podmíněna i vznikem nových, vhodných projekčních postupů a metod. Ukazuje se, že jednou z cest je předřazení datové analýzy a syntézy před podrobnou procesní analýzou jednotlivých úloh systému. Zde naznačená metoda čtyřfázového postupu je nástroj, který umožňuje strukturovat data již ve fázi poměrně hrubého funkčního popisu aplikací. Navíc, ve shodě se současným trendem, zobrazuje bázi dat ve třech úrovních :

- uživatelské, strukturované podle informačního obsahu dat, bez závislosti na počítačových prostředcích a tedy srozumitelné počítačovým laikům,
- programátorské, reprezentované prostředky popisu dat daného systému řízení báze dat /SRBD/ nebo logického IOCS,
- počítačové, odpovídající fyzické realizaci na počítači.

Čtyřfázová metoda určuje jednotlivé fáze a v rámci nich kroky procesu tvorby báze dat informačního systému. Popisuje způsob jakým lze vytvořit použitelnou, dalšího rozvoje schopnou, počítačově realizovatelnou bázi dat.

Zahrnuje:

- Analýzu datového fondu, t.j. především jednoznačnou specifikaci informačního obsahu datových prvků a definování vzájemných vztahů datových prvků založených na jejich informačním obsahu.
- Datovou syntézu, umožňující seskupování datových prvků do datových skupin na základě informačního obsahu datových prvků a současně vedoucí k zachycení vazeb mezi datovými skupinami. Výsledkem datové syntézy je datový model, který obsahuje všechny definované datové prvky datového fondu.
- Popis datového modelu nebo jeho částí ve formě srozumitelné programátorům a počítačovým prostředkům, jenž respektuje zpracovací požadavky na data a omezení dané současnými programovými a technickými prostředky, tzv. logickou datovou strukturu.
- Počítačové řešení báze dat, tzv. fyzickou datovou strukturu.

2. Popis fází

Čtyřfázový postup tvorby báze dat lze rozdělit do čtyř fází, které zachycují jednotlivá stádia tvorby báze. Pro přehlednost jsou první a druhá fáze ještě detailněji členěny na kroky. Každá fáze je uzavřena vytvořením jednoho nebo více výstupů. Fáze jsou nazvány takto:

- datová analýza,
- datová syntéza,
- transformace do logické datové struktury,
- transformace do fyzické datové struktury.

2.1 Datová analýza

Analýza dat v rámci čtyřfázového postupu zahrnuje rozbor a inventuru datových prvků a analýzu vztahů mezi datovými prvky. Analýza se provádí ve dvou krocích.

Náplní prvého kroku je rozbor datových prvků informačního systému, jehož cílem je jednoznačná specifikace datových prvků. Datové prvky se analyzují z hlediska vhodnosti pojmenování, možnosti jednoznačného stanovení informačního obsahu, formátu, jedinečnosti, vztahu k ostatním datovým prvkům, t.j. zda se jedná o datový prvek prvotní / t.j. takový, který je třeba nezbytně v bázi dat uložit /, nebo o datový prvek, který lze z prvotních datových prvků určit.

Výsledkem tohoto prvního kroku datové analýzy je především systematický popis datových prvků, alespoň v rozsahu minimální specifikace charakteristik katalogu dat, který by měl zahrnovat tyto jejich charakteristiky: název /označení, jméno, identifikátor /, formát / předběžnou představu o formátu /, stručnou formulaci informačního obsahu / smyslu /, případně též jeho strukturu.

Uspořádání datového fondu informačního systému, jehož podstatným výsledkem je katalogizace dat, je předpokladem druhého kroku analýzy dat, které už přímo vytváří specificky zaměřené podklady pro následnou datovou syntézu čtyřfázového postupu. Vychází se přitom z informačního obsahu datových prvků a ze vzájemné informační souvislosti datových prvků.

Hlavní náplní tohoto druhého kroku je analýza vztahů mezi základními a přídatnými datovými prvky. Analýzu je vhodné provádět metodou expertního odhadu, kdy se ke každému datovému prvku datového fondu určují jeden nebo více základních datových prvků. Výsledkem je přiřazený datový prvek - základní datové prvky. U rozsáhlých informačních syst. může být vhodné jako pomůcku pro odhad přiřazení připravit nejprve seznam datových prvků identifikačního charakteru, jako kandidátů na funkci základních datových prvků. Podle provedení přiřazení lze sestavit předběžný seznam základních datových prvků, jehož rozbor může být podnětem k prohloubení datové analýzy /prvního kroku/ dříve, než se přejde ke druhé fázi tvorby báze dat, k syntéze.

Za hlavní produkty fáze datové analýzy lze považovat:

- katalog dat,
- přiřazený datový prvek ~ základní datové prvky,
- předběžný seznam základních datových prvků.

Činnosti první fáze vycházející z obsahové analýzy dat metodou expertního odhadu prokazují, že datový fond lze rozdělit na dvě části. Na zpravidla méně četnou část - základní datové prvky, kam zařadíme všechny datové prvky, které jsou alespoň jednou definovány jako určující datové prvky. A na početně obvykle četnější část, označovanou jako přídavné datové prvky, kam zařazujeme všechny ostatní datové prvky.

2.2. Datová syntéza

Datovou syntézou se ve čtyřfázovém postupu rozumí proces tvorby datového modelu. Vychází se přitom z přirozené schopnosti dat seskupovat se na základě informačního obsahu.

Pro rozdělení datových prvků do skupin se zdá být nejvhodnější využití shlukovacího postupu, kdy kritériem pro tvorbu shluků / datových skupin / je shodnost základních datových prvků, které byly v předchozí fázi každému datovému prvku přiřazeny. To znamená, že do jedné skupiny se zařadí všechny datové prvky, které jsou spojeny se stejnou kombinací základních datových prvků. Toto rozdělení datových prvků do skupin je první krok při tvorbě datového modelu. Jeho výsledkem je seznam datových skupin, kdy pro každou datovou skupinu jsou uvedeny základní datové prvky, jež ji identifikují a ostatní datové prvky, které jsou do ní zařazeny. Tento seznam lze považovat za výchozí specifikaci datového modelu.

Poté začíná vlastní modelovací činnost, kterou lze označit za druhý krok datové syntézy. Projektant postupně seznam upravuje a jednotlivé přijaté / upravené / datové skupiny identifikuje číslem a pojmenováním. Úpravy vedou k potvrzení nebo změně jak identifikačních základních datových prvků, tak ostatních da-

tových prvků skupiny; mohou vést i ke sloučení nebo naopak k rozdělení datových skupin stanovených v prvním kroku datové syntézy. Nutnost úpravy výchozí verze není dána jenom tím, že ve fázi datové analýzy nelze očekávat úplnou a bezchybnou specifikaci výchozího přiřazení datového prvku - základní datové prvky, ale i tím že je třeba datový model upravit ve smyslu dále uvedené metody, kterou označujeme jako přiřazovací analýzu. Teprve pak lze očekávat, že datový model bude mít očekávané vlastnosti.

Výchozí verze datového modelu je dána poměrně hrubým a někdy též ne zcela přesným vzájemným přiřazením datových prvků. Úkolem použití přiřazovací analýzy, jejíž základním metodickým nástrojem je expertní odhad, je modifikace výchozí verze datového modelu do formy, která odpovídá základní zásadě tvorby datového modelu a to sdružování datových prvků do homogenních, ucelených skupin podle informačního obsahu a bezredundantního umístění přídavných datových prvků. Při aplikaci přiřazovací analýzy se postupuje podle těchto bodů:

1. Posouzení homogenity a úplnosti každé datové skupiny.
2. Posouzení datových skupin s obdobnými seskupeními základních datových prvků.
3. Posouzení opakovaného umístění přídavných datových prvků.

Třetím krokem tvorby datového modelu je zobrazení jeho struktury ve formě síťového grafu. Prvky grafu jsou základní datové prvky a datové skupiny, vazby jsou dány přiřazením základních datových prvků datovým skupinám, přičemž jsou to právě základní datové prvky, které jsou sprostředkovateli vazeb mezi datovými skupinami.

V souhrnu lze návrh datového modelu označit jako proces uspořádání datových prvků do celkové struktury reprezentující zobrazení vztahů mezi datovými prvky. Datový model zachycuje dvě úrovně vztahů, a to :

- vztahy základní datové prvky - ostatní datové prvky příslušné

datové skupiny,

- vztahy mezi datovými skupinami, jež jsou zprostředkovány základními datovými prvky.

2.3. Transformace do logické datové struktury

Datový model je podkladem pro prvou i všechny další verze logické datové struktury. Logickou datovou strukturou se rozumí modifikace buď celého nebo části datového modelu, kterou lze popsat prostředky popisu dat pro počítačové zpracování.

Tato modifikace je vynucena současnou omezeností počítačových prostředků / technických i programových /, které nejsou schopny bez úprav efektivně realizovat obecnou síťovou strukturu datového modelu. Je třeba zpravidla velmi pečlivě zvažovat vhodnost struktury / především počet vazeb / a také zpracovací požadavky / požadovaný způsob přístupu k daným datům, předpokládanou četnost přístupů, četnost výskytů jednotlivých záznamů, rozsah zpracovávaných dat při jednom přístupu, časové limity odezvy, atd. /.

Jedná se o fázi, která je vynucena tím, že zatím nejsou běžně k dispozici počítačové / programová i technické / prostředky, které by byly schopny bez úprav a efektivně realizovat obecnou síťovou strukturu datového modelu, navíc nezávisle na zpracovacích požadavcích.

Logická datová struktura je tedy výsledkem transformace datového modelu do takové datové struktury, která je popisatelná pomocí jazyka pro popis dat daných programových prostředků.

2.4. Transformace do fyzické datové struktury

Fyzická datová struktura je označení pro formu, jakou je realizována logická datová struktura na počítačovém systému.

Transformace logické datové struktury do fyzické datové struktury je již běžně zvládnutá záležitost. Provádí se automatizovaně na základě specifikace v jazyku popisu dat pomocí programových prostředků. Všechny dnes používané počítačové systémy takové prostředky mají.

3. Počítačová podpora čtyřfázového postupu

Tvorba báze dat je proces, který je determinován mnoha různými, často protichůdnými hledisky. Z praktického hlediska je největším problémem velký počet datových prvků a velký počet vazeb mezi nimi. Datových prvků bývají sta či tisíce a snadno lze nalézt informační systémy disponující ještě vyšším počtem datových prvků. Vazeb bývá 5-15 krát více než datových prvků. To jsou počty, které bez pomoci počítačového zpracování nelze zvládnout. A mimo tato výchozí data je třeba evidovat jak mezivýsledky, tak i konečnou strukturu báze dat.

Proto je i čtyřfázový postup podporován počítačovou evidencí, zpracováním i prezentací dat. Počítačovým prostředím je výpočetní zařízení ADT 4500 v běžné konfiguraci, vybavené operačním systémem DOS - IV nebo RTS / RTE / a systémem řízení báze dat IMAGE 1000.

Programová podpora čtyřfázového postupu se nazývá DISY2 a je tvořena těmito částmi:

- a/ datovou strukturou,
- b/ sadou ukládacích programů,
- c/ dialogovým programem MAN,
- d/ sadou výběrových a tiskových programů pro údaje uložené v databázi DISY2,
- e/ sadou programů pro tvorbu katalogu dat,
- f/ sadou programů ANA pro interakční tvorbu báze dat čtyřfázovým postupem.

Programy ad b/-f/ jsou psány ve FORTRANU4X a využívají podprogramů knihovny KSPD, u výběrových programů se využívá též možnosti subsystému DOTAZ /QUERY/ - IMAGE.

ad a/ Datová struktura je základem DISY2 a umožňuje uložení všech údajů a jejich vazeb potřebných pro tvorbu báze dat čtyřfázovým postupem. Dovoluje evidovat:

- datové prvky,
- úlohy,
- soubory,
- datové skupiny,
- základní datové prvky /klíče/,
- popisy datových prvků

a vazby:

- datové prvky - úloha,
- datové prvky - základní datové prvky /klíče/,
- datové prvky - popisy datových prvků,
- datové prvky - soubory,
- klíče - soubory,
- základní datové prvky - datové skupiny.

Datová struktura DISY2 je koncipována jako databáze tvořená 12 druhy záznamů /setů/, řízená systémem řízení báze dat IMAGE 1000. Obsahuje všechny údaje potřebné pro podporu čtyřfázového postupu a uložení výsledků jednotlivých fází.

ad b/ Při prvotním naplňování záznamů obsahujících údaje o datových prvcích a jejich popisech, o úlohách informačního systému, o vazbách datový prvek - základní datové prvky a o vazbách datový prvek - úlohy, se ukazuje jako užitečné využívat speciálních ukládacích programů. Jedná se zpravidla o velké objemy dat a interakční ukládání dat pomocí programu NAM nebo prostředků řízení báze dat IMAGE není v takové situaci efektivní. Navíc data mohou být připravena v útvaru přípravy dat. Ukládací programy umožňují dávkový sběr dat, přičemž zajišťují potřebné formální kontroly vstupních dat a jejich přeformátování do po-

doby vyžadované systémem řízení báse dat IMAGE. Programy jsou dialosové, ukládání probíhá prostřednictvím obrazovky a jsou řešeny tak, že operátora při ukládání vedou.

ad c/ Program NAM je dialogový, segmentovaný program. Pracuje s libovolným záznamem datové struktury DISY2. Dovoluje všechny čtyři potřebné základní typy operací se záznamy:

- ukládání nových záznamů,
- aktualizaci záznamů,
- rušení záznamů,
- čtení záznamů.

Omezení typu operací při práci se záznamy jednotlivých typů je dáno funkcí typů záznamů. U vazbových setů není povolena aktualizace, neboť každá změna je fakticky definicí nové vazby. Dochází-li ke změně, je třeba zrušit existující vazby a uložit vazby nové. U automaticky vytvářených záznamů má smysl pouze čtení.

Protože program NAM je základním prostředkem pro evidenci datového modelování při čtyřfázovém postupu, je vhodné připomenout ve stručnosti fungování programu.

Po spuštění se program NAM ohlásí a nabídne tabulku s označením typů záznamů databázové struktury DISY2. Operátor /projektant/ zvolí typ záznamů, se kterým si přeje pracovat. Program zobrazí strukturu záznamu a povolené typy operací s ním. Operátor zvolí požadovaný typ operace. Obdobným způsobem - v závislosti na typu operace - probíhá komunikace operátor - program dále. Při práci v libovolném typu operace může operátor prakticky kdykoli využít řídicí příkazy, které dovolují ukončení práce programem, přechod na rozhodovací blok pro výběr typu operace, volbu typu operace, návrat na předchozí položku záznamu, přechod na následující položku záznamu.

Program NAM provádí patřičné formální i vztahové kontroly při ukládání, aktualizaci i rušení záznamů. Dojde-li k chybové-

mu záznamu během práce programem, program vypíše hlášení o chybě a podle povahy chyby se buď ukončí sám nebo očekává příslušný řídicí příkaz od operátora /projektanta/.

Programově je zabezpečen i tisk ukládaných, aktualizovaných a zrušených záznamů i změna kódu stavu záznamu.

ad d/ Výběrové a tiskové programy, které jsou součástí DISY2, umožňují výpis údajů libovolného typu záznamu pro definovanou množinu těchto záznamů v seřídění, které lze zadat. Údaje uložené v databázi DISY2 jsou tak projektantovi báze dat kdykoliv k dispozici.

ad e/ Programy tvorby katalogu jsou v podstatě podskupinou výběrových a tiskových programů. Uvádíme je však zvlášť, vzhledem k jejich specifitě, jednoznačnému zaměření i důležité roli, kterou katalog dat hraje v projektování informačních systémů moderními metodami.

ad f/ K nejpoužívanějším programům této sady ANA patří :

- program ANA5 pro seskupení datových prvků podle shodnosti základních datových prvků a tisk těchto seskupení,
- program ANA2 pro výpis datových skupin a to buď vybraných nebo všech; pro každou datovou skupinu obsahuje výpis základních i přídatných datových prvků,
- program ANA3 pro výběr údajů o struktuře vztahů mezi datovými prvky; program vypíše přiřazení základní datové prvky - přídatné datové prvky a pro každý přídatný datový prvek další k němu uvedené základní datové prvky,
- program ANA1 pro vyhledání a výpis datových prvků, které jsou společné pro základní datové prvky zadané v dialogu.

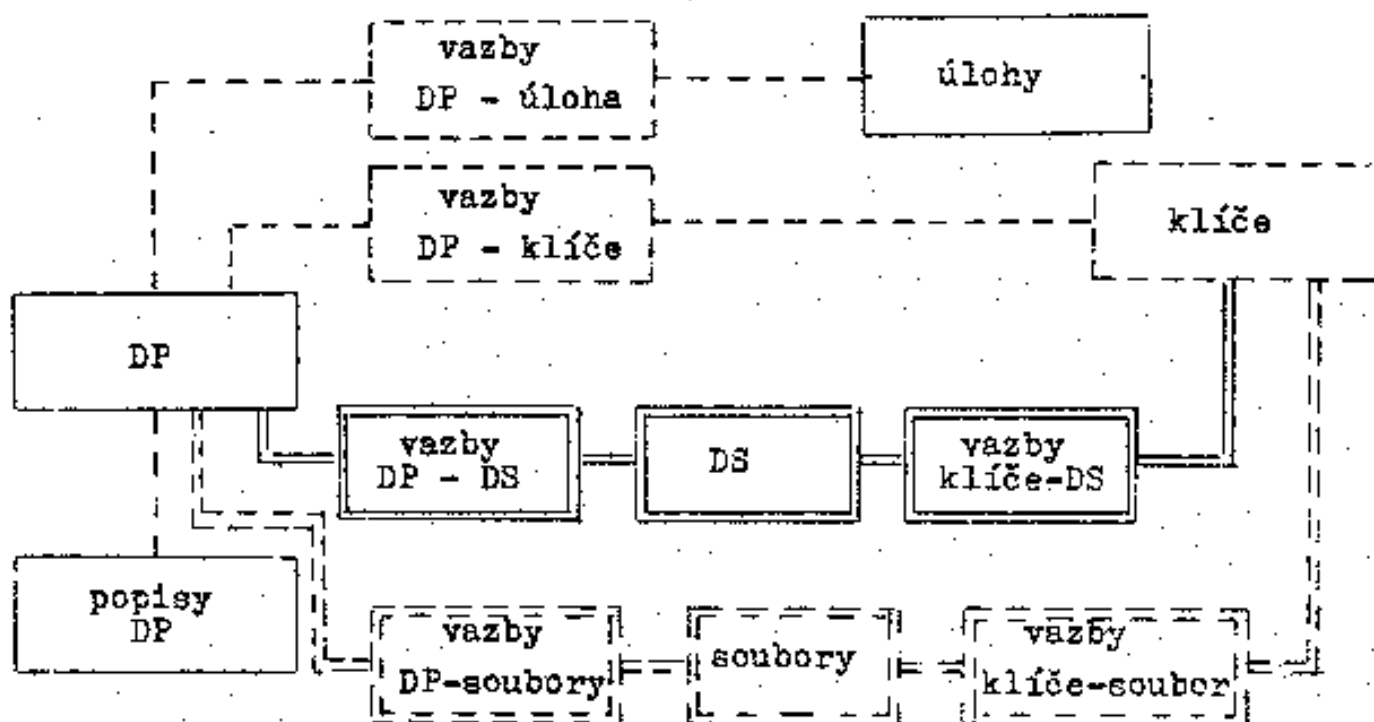
4. Z á v ě r

S rozvojem využití počítačů v informačních systémech se kromě problémů návrhu vhodných algoritmů objevuje i problém, jak správně strukturovat data. Tento příspěvek ukazuje metodu, která tento problém řeší poměrně jednoduchým způsobem. Metoda je prakticky využívána a dále se pracuje na jejím rozvoji, jak ve směru zpodrobnění jejích kroků, tak na inovaci programové podpory DISY2.

Literatura

- /1/ Havlík, J., Zralý M. : DISY2 - Programová podpora projektování ISR/ASR - popis, Praha 1985, nepublikováno
- /2/ Zralý M. a kol. : Tvorba bází dat informačního systému pro řízení, příručka kursu DT ČSVTS Ústí n.L., 1985, s. 82
- /3/ Havlík J., Zralý M. : Prototypový přístup k projektování informačního systému pro řízení, Sešity Inorga č. 115, Praha 1985, s. 85

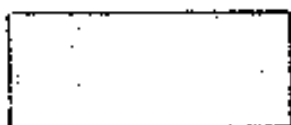
Obrázek struktury databáze:



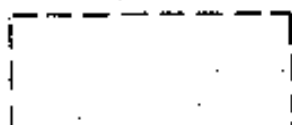
DP datové prvky

DS datové skupiny

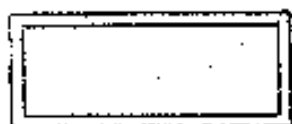
Poznámka: V obrázku jsou graficky odlišeny sady v závislosti na fázi naplňování daty při použití čtyřfázové metody



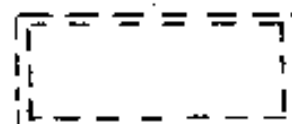
1. prvotní naplnění
sady: úlohy, datové prvky, popisy DP



2. naplnění na základě datové analýzy
sady: vazby DP-úlohy, vazby DP-klíče, klíče



3. naplnění při datové syntéze do dat. modelu:
datové skupiny, vazby DP-DS,
vazby klíče-DS



4. naplňování při odvozování logické datové struktury:
soubory, vazby klíče-soubory,
vazby DP-soubory