

Ing. Vilém Keller

Příspěvek shrnuje dosavadní názory na databanky v ASŘ, podává stručný teoretický nástin topologie datovýchází založené na matematické teorii bodových množin, uvádí některé ekonomické aspekty zpracování dat a odtud činí pokus o hledání odpovědi na otázky po směrech pravděpodobného vývoje programování v databankovém prostředí.

### 1. Veliký problém software

V tomto příspěvku si budeme všimát především aplikačního programového vybavení (APV) pro zpracování dat v řízení na úrovni podniku (ASŘP). Vycházíme z faktu, že naprostá většina kapacit výpočetních systémů a také analyticko-programátorské práce je využívána právě zde. Ostatně zde vznikl i tzv. veliký problém software.

Co rozumíme pod tímto pojmem? Je obecně známo, že hodnota APV zřetelně přesahuje pořizovací cenu počítače. Tento neponěr ještě poroste s tím, jak budou dále klesat ceny. Zatímco technická výkonnost stroje s postupujícím vývojem rychle roste, výkonnost tvůrců APV tento trend sleduje daleko volněji. Byla učiněna řada pokusů o zvýšení produktivity programátorské práce - víme dobře, které to jsou - ale můžeme říci, že přes určité úspěchy zde nebylo dosaženo zdaleka tolik, čeho docílili tvůrci hardware. Pracnost průměrného ASŘP stále zůstává vyšší než 100 pracovních roků a neklesá, spíše roste s tím, jak rostou požadavky na strojové zpracování. Riziko neúspěšných aplikací sice klesá, avšak stále je přítomno a stále hrozí. To se projevuje nejen častými změnami jednotlivých programů, ale i občasným "přeprogramováním" celých skupin úloh. Přidáme-li k tomu ještě vlivy okolí např. měnící se předpisy a metodické pokyny, pak není divu, že významnou část programátorských kapacit je nutno věnovat na údržbu provozovaného APV.

A tak veliký problém software se důrazně projevuje tímto paradoxem: Jakmile dosáhneme stavu, kdy naše APV funguje spoleh-

livě, kdy uživatelé dostávají správné podklady a iniciativně žádají další, kdy také náš provoz vykazuje příznivé hodnoty ukazatelů v ročních výkaze, právě tehdy náš počítač dožívá svůj fyzický věk, musí být vyměněn a celá stavba APV se povážlivě zakymácí. Řekneme-li totéž obráceně, pak platí zákon, že se svými APV začneme být spokojeni právě v době, kdy má být opuštěno.

A tak může kolotoč "člověkoroki" začít znovu. Tentokrát však pod nevidaným časovým tlakem, neboť uživatel si již zvykl, žádá své zde a ihned a není zpravidla podruhé ochoten podstoupit utrpení všech těžkých začátků. Zlaté časy prvního nadšení jsou v ne-návratnu.

Do této situace nyní dramaticky zasahuje idea databanky. Ta ovšem přináší novou kvalitu a problematizuje dřívější koncepce ASŘP založené na dávkovém zpracování tradičních souborů dat. Idea databanky důrazně podporuje argumentaci o efektivnosti počítačů v řízení a nyní se jeví jako nadějný východiště při řešení problému jak dál v ASŘP. Přináší však další problém - jak dál s databankami.

Nežli provedeme další úvahy, povšimněme si tři typických oblastí aplikace databank. Jsou to:

1. Systémy zpracovávající údaje, jejichž typickým příkladem jsou vědeckotechnické informace.
2. Systémy sloužící účelům, jejichž typickou reprezentací je rezervace letenek.
3. Systémy zpracování dat pro řízení (ASŘP).

Každá z nich má svůj charakteristický datový obsah a způsob práce s ním a všechny se navzájem podstatně odlišují. Prvé dvě aplikace jsou velmi dobře metodicky propracovány, ve světě hojně užívány a poskytují mnoho výhod při relativně malé námaze potřebné k jejich implementaci. Ve třetí oblasti však existuje zatím řada dosud nezvládnutých problémů a právě zde není dostatek praxí dobře prověřených aplikací. Zatímco o efektivnosti databank v prvních dvou oblastech sotva kdo pochybuje, o databankách v ASŘ se vedou neustálé diskuse. Pokusíme se je shrnout a pak hledat východiště ze situace, kterou jsme označili jako veliký problém software.

## 2. Nástin topologie datovýchází

Abychom našli společný jazyk pro další úvahy, zavedeme nyní některé pojmy známé z teorie množin a z formální logiky, abychom je pak aplikovali na problematiku struktury datovýchází.

### 2.1 Zobrazení

Nechť je dána množina  $X$  s prvky  $x$  a množina  $Y$  s prvky  $y$ .  
Nechť je dále dán kartézský součin  $K$  množin  $X$  a  $Y$  jako množina všech dvojic  $(x,y)$  v tomto pořadí. Potom zavedeme pojem zobrazení množiny  $X$  do množiny  $Y$  pomocí výrazu

$$(x) (Ey) [f(x,y)] \quad (1)$$

který čteme: ke každému  $x$  existuje alespoň jedno  $y$  takové, že dvojice  $(x,y)$  má vlastnost  $f$ . Tato vlastnost ve smyslu logiky tříd definuje množinu stejného jména, takže můžeme psát, že  $(x,y) \in f$ . Říkáme, že  $y$  je obrazem vzoru  $x$  v zobrazení  $f$ , což můžeme zapsat také  $y = f(x)$ . Odtud  $X$  je množinou vzorů a  $Y$  množinou obrazů. Můžeme také říci, že zobrazení  $X$  do  $Y$  je určitý jednoznačný předpis, který ke každému prvku  $x \in X$  přiřazuje jistý prvek  $y \in Y$ .

Poznámka 1. Prvky množin  $X$ ,  $Y$  mohou být opět množiny, pak se jedná o zobrazení soustavy množin  $\bar{X}$  do  $\bar{Y}$ . Příklad  $\bar{X} = \bar{Y}$  není vyloučen.

Poznámka 2. Každý prvek  $y \in Y$  nemusí být nutně obrazem některého  $x \in X$ . Jestliže však jím je, pak mluvíme o zobrazení  $X$  na  $Y$ .

Poznámka 3. Má-li každý vzor  $x$  jeden a jen jeden obraz  $y$ , a naopak každý obraz  $y$  jeden a jen jeden vzor  $x$ , pak se zobrazení nazývá prosté, někdy též jednojednoznačné. Jinak může mít jeden vzor více obrazů, což můžeme značit  $1 : n$ , nebo jeden obraz více vzorů, což můžeme značit  $m : 1$  anebo platí oboje, což můžeme značit  $m : n$ .

Nechť je  $f$  zobrazení  $X$  do  $Y$  a nechť je  $\bar{X}$  soustava všech podmnožin množiny  $X$  a  $\bar{Y}$  soustava všech podmnožin množiny  $Y$ . Zobrazením  $f$  je určeno zobrazení  $f^1$  soustavy  $\bar{X}$  do  $\bar{Y}$  a zobrazení  $g$  soustavy  $\bar{Y}$  do  $\bar{X}$  takto:

- (A) když  $M \in \bar{X}$ , pak  $f(M)$  je množina obrazů všech  $x \in M$  při zobrazení  $f^1$ ,  
 (B) když  $N \in \bar{Y}$ , pak  $g(N)$  je množina všech těch  $x \in X$ , jejichž obraz  $f(x)$  náleží do  $N$ .

Zobrazení  $f^1$  můžeme bez obav z nedorozumění značit prostě  $f$  a zobrazení  $g$  můžeme značit též  $f^{-1}$ . Jestliže je  $f$  prosté zobrazení  $X$  na  $Y$ , pak zobrazení  $f^{-1}$  nazýváme inverzním zobrazením k  $f$ .

Nechť jsou dány množiny  $X, Y, Z$  a zobrazení  $f$   $X$  do  $Y$  a zobrazení  $g$   $Y$  do  $Z$ . Položíme-li

$$h(x) = g(f(x)) \quad \text{pro každé } x \in X \quad (2)$$

dostaneme zobrazení  $h$  množiny  $X$  do množiny  $Z$ , o kterém říkáme, že je složeno ze zobrazení  $f$  a  $g$  (pořadí je podstatné!). Jinak řečeno,  $h$  je složené zobrazení  $X$  do  $Z$ .

Pojem zobrazení má v našich úvahách základní význam. Kromě něho se užívá též pojmů transformace, projekce, funkce, operace, korespondence atd. Zde se přidržujeme původní terminologie našich matematiků zejména lit. [1].

## 2.2 Uspořádaná báze

Nechť je dána (diskrétní konečná neprázdná) množina  $P$  a soustava  $\bar{X}$  jejích podmnožin. Nechť je dále dána množina zobrazení  $F$  v soustavě  $\bar{X}$  mající tuto vlastnost:

- (C) Jednotlivé prvky soustavy  $\bar{X}$  lze uspořádat do posloupnosti  $R(X)$  tak, že pro každý prvek  $X$  platí, že neobsahuje obrazy žádného prvku, který jej v posloupnosti následuje.

Z toho plyne, že posloupnost  $R(X)$  má alespoň jeden počáteční a alespoň jeden konečný člen. Pak soustavu  $\bar{X}$  spolu s množinou zobrazení  $F$  nazveme uspořádanou bází.

## 2.3 Graf báze

Grafem báze nazveme takový graf, jehož uzly jsou tvořeny množinami soustavy  $\bar{X}$  a jeho hrany zobrazeními  $f \in F$ . Z vlastností uspořádané báze vyplývá, že je její graf acyklický.

## 2.4 Úplná báze

Graf uspořádané báze zobrazíme maticí  $M$ , v níž

$m_{ij} = 1$  existuje-li zobrazení množiny  $X_i$  do  $X_j$

$m_{ij} = 0$  jinak.

Vytvářejme nyní posloupnost booleovských součtů booleovských mocnin matice  $M$  podle vztahu

$$S_i = \sum_{j=1}^n M^j \quad (3)$$

Pro jisté  $i \leq n$  může horní trojúhelníková submatice matice  $S_i$  nabýt vesměs jedničkových prvků. V kladném případě budeme mluvit o úplné bázi.

V úplné bázi je možné nalézt složené zobrazení nejvýše  $n$ -tého řádu každé množiny do každé jiné, která jí v posloupnosti následuje. Přiblížneme-li k vlastnostem zobrazení  $f^{-1}$ , pak také do každé, která jí v posloupnosti předchází.

## 3. Řízení a informace

### 3.1 Systém řízení

Na reálném objektu "podnik" lze definovat velmi veliké množství systémů. Ba i sám systém řízení může být definován ve velmi mnoha podobách. Počítač však neumí více (ale také méně) než zpracovávat data, a přiblížneme-li k tomu, že program je procedurou algoritmického charakteru, a že zatím více neumíme, pak nám nezbyvá, než omezit se na zcela formalizovatelné systémy, jejichž chování lze popsat prostředky počítačového programu. Pak se nám systém řízení bude nutně jevit jako systém (definovaných) činností v organizační struktuře, mezi nimiž existují informační vazby. A to je již silný stupeň abstrakce.

Informační vazby jsou realizovány jako proces zpracování a předávání zpráv a jsme si vědomi rozdílů mezi pojmy správa a informace. Prvky zpráv jsou data. Jednotné číslo zde zní ovšem dosti nezvykle, a proto je nahradíme slovem údaj. Za údaj budeme

považovat proměnnou veličinu, která může být vyjádřena číslem, otevřenou řečí nebo smluveným kódem nebo může jít o proměnnou logického typu. Každý údaj nabývá hodnot ze svého definičního oboru.

Pozorujeme-li takto systém řízení, pak lze na něm definovat jeho informační podsystém, který označíme jako informační systém pro řízení (ISR). Vybranou část ISR přejímáme do počítačového zpracování a obvykle mluvíme o ASŘ.

### 3.2 Frekvence zpráv

Ve formalizované části ISR je základním nositelem zprávy písemný doklad, který vzniká zpracováním údajů podle předepsaného způsobu. Běžné ASŘP přejímají tento proces, transformují jej a převádějí tak "ruční" zpracování do APV. Tím postupně vzniká ASŘP tradičního typu. Tzv. systémový přístup tento proces uspořádává podle svých pravidel a vnáší sem jistý řád.

APV je tvořeno velkým počtem počítačových úloh. Pozorujme nyní počet průchodů každé úlohy počítačem za celou dobu jeho životnosti a seřadme tyto úlohy podle tohoto počtu. Vynesením četnosti do grafu získáme tzv. Lorenzovu křivku, která vypovídá, že cca 10 - 20 % nejčetnějších úloh se podílí 80 - 90-ti procenty na všech průchodech počítačem a ovšem naopak zbývajících 80 - 90 % úloh se podílí jen 10 - 20-ti procenty na všech průchodech. Tento jev je znám v ekonomii a souvisí s tzv. Paretovým rozdělením pravděpodobnosti. Stejným zákonitostem ovšem podléhají i četnosti zpráv v ISR vůbec.

Odtud plyne, že 10 - 20 % programátorské práce v tradičním APV pokrývá 80 - 90 % vytižení počítače a také výstupů do systému řízení podniku. Číslo se mohou případ od případu poněkud lišit, nicméně hlavní rys této zákonitosti, to jest veliká nerovnoměrnost rozdělení četností, zůstává zachován.

### 3.3 Báze dat

Báze dat tradičního ASŘP je tvořena soubory dnes již ustáleného typu. V jazyce kapitoly 2 je to soustava  $\bar{X}$  množin údajů  $X$ . Sjedením všech množin  $X$  vznikne množina  $P$  jednotlivých údajů  $p$ .

Soustava  $\bar{X}$  není disjunktní, tzn. jeden a týž údaj  $p$  se může vyskytovat ve více množinách  $X$  a pak mluvíme o redundanci. Ta se tedy projevuje tak, že součet všech prvků soustavy  $\bar{X}$  je větší číslo než počet prvků množiny  $P$ .

Množina  $F$  zobrazení  $\bar{X}$  do  $\bar{X}$  zde vyjadřuje vztahy mezi soubory. Ty jsou povětšinou implicitní a to zde znamená asi tolik, jako bychom řekli, že analytikům stačí, když o nich vědí a že programátoři je respektují. Explicitně jsou některá zobrazení dána jen ve vybraných případech, ponejvíce tam, kde je nutno brát ohledy na souvislosti údajů uvnitř subsystémů a někdy i mezi nimi navzájem.

Databankový přístup - zjednodušeně řečeno - redukuje soubory do normálních forem a vytváří domény. Soustava domén  $\bar{X}$  je podmíněně disjunktní. Množina zobrazení  $F$  soustavy  $\bar{X}$  do  $\bar{X}$  zde má zásadní význam a musí být explicitně dána. Domény obvykle vytvářejí uspořádanou (a úplnou) bázi. Je-li  $P$  množina všech prvotních údajů, pak musí existovat ještě množina  $Q$  obsahující údaje, které popisují každé zobrazení  $f \in F$ . Zrušení redundance prvotních dat je tak zčásti nebo i zcela zapláceno vznikem množiny  $Q$ , jejíž prvky nesou informace nutné k nalezení údajů logicky k sobě vázaných, umístěných však v různých doménách.

### 3.4 Databáze a frekvence zpráv

Pozorujeme chování tradičního a databankového systému v provozu. Tradiční soubor čteme prostě větu za větou sekvenčně nebo přímo podle potřeby a zpracování organizace souboru. Žádáme-li z databanky větu zadané struktury, čteme ve všech doménách, v nichž se jednotlivé údaje vyskytují, a to za podmínky, že existují potřebná zobrazení. Tato transakce je za jiných stejných okolností časově výrazně delší a diference roste s počtem údajů ve větě.

Snadno se přesvědčíme o tom, že práce s tradičním souborem při vysokých četnostech průchodů dloh bude významně efektivnější nežli s databankou. Naproti tomu v té části Lorenzovy křivky, kde 80 - 90 % dloh, které spotřebovaly 80 - 90 % programátorských kapacit, ale vykazují jen 10 - 20 % průchodů, tam slibuje databanka velmi dobré výsledky. Nutnou podmínkou je ovšem existence úplné báze.

#### 4. ... a jak dál

##### 4.1 Shrnutí diskuse o databankách

Kolem problematiky spojené s databankami je vedena čilá diskuse. To ovšem věci nikterak neubírá na významu, spíše naopak. Pokusíme se velmi stručně vybrat to, co se považuje za výhody a to, co se považuje za nároky nebo i nevýhody.

Nejprve výhody:

1. Odstranění redundantních údajů.
2. Zabezpečení obsahu proti neoprávněnému přístupu.
3. Nezávislost APV na datech.
4. Možnost současného přístupu více uživatelů k datům ve stejném čase, a to nezávisle na programátorském zprostředkování.
5. Racionalizace ředitelské práce řádově o stovky procent.
6. Snížení objemu děrování, práce vstupní kontroly a množství potiskového papíru.
7. Jsou k dispozici vždy aktuální data.
8. Racionalizace rozhodovacích procesů.
9. Zvýšení pořádku v datech.

Naproti tomu nutno vzít v úvahu tyto faktory:

1. Nutnost zvýšení úrovně a preciznosti analýzy a programování.
2. Podstatně vyšší nároky na výkonnost a spolehlivost hardware.
3. Nutnost zřídít náročné pracoviště správy dat.
4. Riziko vyvolání rozsáhlých změn v dosavadní organizační struktuře, informačním systému a v hierarchii řídicích úrovní.
5. Riziko zahlcení stroje při rostoucím rozsahu báze dat a tím i riziko selhání celého systému.
6. Zpomalení aplikačních programů vlivem nutného procesu přípravy dat prostřednictvím SŘBD.
7. Nedostatek disponibilních standardních SŘBD pro převládající typ počítačů u nás.

Uvádí se, že v současné době u nás existuje něco více než 100 aplikací databank. Jestliže pomíneme ty z nich, které dokazují nejvýše tolik, že to přes databanku jde také, pak zjistíme, že mají tato společná:

1. Struktura databáze je v dobré shodě s ISŘ.
2. Nad rozsáhlými databázemi se většinou operuje vlastním APV a standardní SŘBD se neužívá.



3. Standardní SŘBD se úspěšně používá pro menší a malé rozsahy databáze.
4. V případech aplikace standardního SŘBD nad malou bází dat dochází k prokazatelné úspoře programátorské práce oproti tradičnímu přístupu, i když úspory řádově o stovky procent nebyly zjištěny.

Stojí za to poznamenat, že při pozornějším porovnání diskuse s praktickou aplikací nám neujde, že databanky byly úspěšné až velmi úspěšné tam, kde šlo o aplikaci dosti podobnou prvním dvěma oblastem uvedeným v kapitole 1.

#### 4.2 Databanka a velký problém software

To, co jsme označili za velký problém software, má tyto důsledky:

- s tím, jak roste objem AFV, zpomaluje se tempo vývoje vlivem ztráty kapacit na údržbu již realizované části;
- postupný vývoj AFV způsobí, že po větši část doby vývoje není ASŘP systémem, i když projekt sám je jako systém založen;
- za dobu životnosti prvního počítače nelze vybudovat ASŘP do celé šířky i hloubky tak, aby jeho přínos pro zvýšení úrovně řízení a tím i ekonomické efektivity podniku byl naprosto bezesporný;
- přechod na nový počítač u tradičního AFV bývá často přirovnáván k živelné pohromě; u databanky, kde není zajištěna portabilita SŘBD, se dá mluvit o katastrofě.

Idea databanky by zde mohla svádět k domněnce, že postačí přestavba tradičních souborů a aplikace standardního SŘBD k uspokojivému vyřešení velkého problému software. Nutno však konstatovat, že veliký problém software bude trvat dotud, dokud praxe nepotvrdí alespoň tu diskusní tézi, která mluví o racionalizaci říditelské práce řádově o stovky procent.

#### 4.3 Přijatelná východiška

Přes všechny problémy však platí, že idea databanky zakládá hlavní směr vývoje počítačového zpracování dat pro řízení. Do doby, kdy budou naléhavé problémy vyřešeny, budeme ještě nuceni vycházet ze svého dosavadního AFV a přizpůsobovat je vývojovým trendům. Postupně bude možno zúžit okruh prací na oněch 10 - 20 % nej-

četnějších úloh a zbytek rozdělit mezi další tvorbu APV a prostředky, které nám poskytnou standardní SRBD. Databanky relačního typu se zde jeví jako velmi nadějně.

#### 4.4 Co lze očekávat

O budoucnosti víme jen tolik, že určitě nějaká přijde a že bude v mnohem jiná, než dnes očekáváme, a téměř jiná, než si přejeme. Přesto se pokusme odvodit některé tendence možného vývoje na základě toho, co víme dnes.

Očekává se, že dojde k patrným změnám dělby práce v naší programátorské profesi. Programátor zaměřený převážně na stroj a algoritmus se daleko zřetelněji oddělí od projektanta - analytika a bude se věnovat vývoji účelově zaměřeného software, a tak se odkloní od tvorby APV. Vzroste význam profese projektanta ISŘ a náplň i styl jeho práce se výrazně odliši od činnosti dnešního analytika - programátora. Kdo si řekl, že počítače mohou významně zlepšit řízení dobré, ale že pouze prodraží řízení špatné. V tom je, zdá se, klíč k pochopení jeho úlohy.

Pod vlivem nového zaměření výuky poroste podíl pracovníků v řízení a správě, kteří si se znalostmi přinesou i pozitivní postoje k denní práci s počítačem a užívání problémově orientovaných jazyků se postupně stane přirozeností.

Naznačený vývoj nebude zdaleka plynulý a bezrozporný. Musíme očekávat vznik celé řady potíží a problémů. Efektivnost počítačů v řízení není pojem, který by byl již zcela vyjasněn a shodně chápán jak uživatelskou sférou, tak i počítačovými specialisty. Idea databanky se časem zcela prosadí, i když struktura báze dat i SRBD mohou doznat podstatných změn oproti tomu, co máme dnes k dispozici.

#### Literatura

- 1 Čech E.; Topologické prostory; Praha, NČSAV, 1960.