

PARAM II - SOUSTAVA PARAMETRICKÝCH PROGRAMŮ PRO NĚKTERÉ ÚLOHY HROMADNÉHO ZPRACOVÁNÍ DAT

Ing. Andrej Boldiš; ~~Dr.~~ Dr. Jaroslav Lizner, CSc.

Popisovaná soustava generátorů označovaných jako PARAM II obsahuje řadu originálních prvků a pojetí, jako např. standardizace operací na větách a souborech, superpozice procedur (funkcí), a jiné. Používání PARAM II má řadu kladných dopadů na projekčně implementační přípravu informačního a řídicího systému a jeho využívání.

1. Úvod

Soustava generátorů PARAM II byla navržena a realizována a v současnosti je používána na Federálním ministerstvu financí Praha. Hostitelským operačním systémem PARAM II je BS 2000 verze 6 firmy Siemens AG. Účelem tohoto referátu je stručně popsat základní rysy tohoto prostředku pro pohotovou a levnou implementaci systému pro zpracování dat.

2. Co není v PARAM II nového

Koncepce generátoru jako prostředku pro zefektivnění implementace systémů na zpracování dat není nová. Nejznámějším a nejpoužívanějším je snad SORT/MERGE generátor. Další obecně uživatelsky zaměřené prostředky bývají součástí základního firemního software (RPG, konverzní programy, a pod.) nebo součástí systémů pro řízení a využívání databází (např. CULPRIT) a pod. Mnohá uživatelská pracoviště si vyvíjí vlastní obecně použitelné softwareové pomůcky. Řekneme si proto hned v úvodu, co není v PARAM II nového.

Generátory obvykle umožňují pracovat s proměnnými operandy, přesněji s libovolnými soubory, větami a s libovolnou strukturou vět. Soubor bývá chápán jako logická jednotka nezávislá na mediu, organizaci vět v souboru a přístupu k větám. Pohotovost a efektivnost bývá zajištěna úspornou formou specifikace funkcí generátoru ve tvaru tabulky nebo formuláře. Každý generátor si zajistí syntaktickou kontrolu parametrů a jejich obsah obvykle převádí do své interní notace.

3. Co je v PARAM II původního

a) PARAM II je pojat jako soustava obecných souborových funkcí s transparentní sémantikou.

b) Soubor nejenže je chápán jako logická jednotka, soubor je chápán i jako hodnota souborové funkce, což je využito v principu superpozice (vkládání) souborových funkcí.

c) Sémantická průhlednost prostých i složených souborových funkcí je zajištěna konceptem tzv. "situací", což není nic jiného než specifická místa v jednotlivých algoritmech souborových funkcí, kde si uživatel musí resp. může předesat procedury na větách.

d) Operace na větách jsou vysoce standardizovány, přitom ale dostatečně obecné.

e) PARAM II je vlastně jakýsi dvouúrovňový procedurální jazyk (viz dále specifikaci funkcí).

f) Je oddělena syntaktická analýza zdrojového textu parametrů a jejich převod do interního zápisu od provádění.

g) Pro ulehčení práce slouží knihovny standardů vět a tisků, tiskových textů a rastrů.

V koncepcích generátorů často se promítají bezprostřední uživatelské potřeby a funkce (např. aktualizace souboru, kontrola souboru, a pod.), což vede k nedostatečné obecnosti. U jiných generátorů se zase promítá víra autorů, že existuje obecný algoritmus pro práci se soubory takový, že každá konkrétní potřeba je pouze speciálním případem tohoto obecného schématu. Takováto snaha o přehnanou obecnost vede k sémantické mlhavosti funkcí programu, nemluvě o tom, že v zpracování dat existují úlohy a algoritmy, které ani zdánlivě nemají nic společného.

Naší snahou bylo a je vypořádat se s řadou problémů, zejména s otázkou produktivity práce při implementaci a efektivnosti výpočtů. Za nejobtížnější a současně nejdůležitější (i když ne právě bezprostřední) ovšem považujeme rozporný problém obecnosti a sémantické průhlednosti funkcí programů. Obtížnost tohoto problému a dopad jeho řešení lze formulovat skutečně pouze mlhavě. V tomto smyslu tvrzení ad a) nutno chápat pouze jako doměnkou autorů.

4. Programy tvořící výpočtové jádro PARAM II

Pro popis funkcí programů použijeme funkcionální zápis, kde vstupní soubor bude vystupovat jako argument funkce a výstupní soubor se bude uvádět jako levá strana příkazu. Příklad: $Z := \text{SORT}(X)$, $Z := \text{SORT}(X_1, X_2, X_3, \dots)$.

Protože většina programů PARAM II realizuje složené funkce, uvedeme si nej-
dříve elementární funkce zabudované v těchto programech. Označení elementár-
ních funkcí je čerpáno z pojednání o relačních datových modulech a z matemati-
ky. Z prostorových důvodů nemůžeme blíže specifikovat jednotlivé funkce.

Z: = INJ (X ₁ , X ₂ , ...)	injekce, (výběr a konverze), na nižší úrovni se uvádí specifikační predikát pro výběr z X a/nebo konstruk- tivní operace
Z: = SEL (X)	select (výběr z databanky)
Z: = JOIN (X, Y)	spojení X a Y
Z: = PROJ (X)	projekce (např. asociativní, statistické)
Z: = MPROJ (X)	multi-projekce (např. statistické)
Z ₁ , Z ₂ , ...: = DECO (X)	dekompozice (rozklad)
Z: = SORT (X)	setřídění
Z: = MERGE (X, Y)	zatřídění
Z: = MNO (X, Y)	konstruktivní množinové operace sjednocení, průniků, rozdílů, symetrických rozdílů, a/nebo rozhodovací o- perace o vztahu náležitosti nebo rovnosti. Druh operací plyne z obsahu procedury na nižší úrovni.

Prozatimní jádro PARAM II tvoří následující programy :

KONS: Z ₁ [, Z ₂]: = DECO (INJ (X ₁ , X ₂ , X ₃ , ...))	
KONJ: Z ₁ [, Z ₂]: = DECO (INJ (X ₁ , X ₂ , ...))	(úspornější KONS)
UPDA: X := INJ (X)	(update)
MNOJ: Z ₁ [, Z ₂]: = DECO (MNO (INJ (X), INJ (Y)))	
MNOD: Z ₁ [, Z ₂]: = DECO (JOIN (INJ (X), INJ (Y)))	
ZATR: Z ₁ [, Z ₂]: = DECO (MERGE (INJ (X), INJ (Y)))	
KOMP: Z := PROJ (INJ (X))	(asociativní proj.)
DEKO: Z := PROJ (X)	(Inverzní ke KOMP)
ADAG: Z ₁ [, Z ₂]: = DECO (MPROJ (INJ (X)))	(statistické proj.)
TRID: Z ₁ [, Z ₂]: = ϕ_1 (SORT (ϕ_2)), ϕ_1, ϕ_2 je některá předcházející funkce s určitým omezením a zjednoduše- ním.	

Obecně ještě platí, že Z₂ (a též DECO) není povinné a argumentem INJ může
být funkce SEL.

Příklad : Z: = MPROJ (SORT (MERGE (INJ (X), INJ (SEL (Y))))
Z: = MNO (INJ (X), INJ (SEL (Y)))

5. Specifikace funkcí programů PARAM II

Funkce programů se specifikují pomocí parametrů. Jednotkou parametrů je formule. Formule jsou členěny do sémanticky příbuzných tříd:

- Formule řídicí, specifikují jméno programu, vstupní, výstupní soubory, pracovní oblasti a řídicí klíče.
- Formule přesunové umožňují předepisovat přesuny mezi větami.
- Formule výpočtové slouží pro vyjádření aritmetických a matematických funkcí.
- Formule transformací slouží speciálně pro vytváření asociativních projekcí.
- Formule typu "search" realizují tabulkové funkce.
- Formule kontrolní umožňující zadávat testy.
- Formule pro řízení procedur umožňující realizovat podmíněné a nepodmíněné skoky, úplné a neúplné cykly.

Jak už bylo řečeno soustava programů PARAM II tvoří jakýsi dvouúrovňový procedurální jazyk. Formule ad a) definují vyšší úroveň, ostatní formule se vztahují k nižší úrovni. Formule ad b) až g) jsou seskupovány podle tzv. "situací" specifických pro algoritmy jednotlivých programů. Např. v programu KONS je předepsána 1 povinná "situace". V programu MNOJ je takovýchto "situací" 5: $[A,] [B,] 1, 2, 3$, tedy A a B nepovinné, ostatní povinné. Nevyužití příležitosti uvést operace v nepovinných situacích vede k speciálním případům jednodušších funkcí programů. Např. neuvědeme-li pro program MNOJ žádnou formuli pro situaci A a B a současně uvedeme-li pouze 1 výstupní soubor, funkce programu se redukuje na $Z := MNO(X, Y)$.

Uvedeme ještě pro ilustraci schema jedné značně obecné výpočtové formule: $z := ((\sum_{i=1}^n x) / (\sum_{i=1}^n y)) \cdot konst1 + ((\sum_{i=1}^n u) / (\sum_{i=1}^n w)) \cdot konst2$, přičemž sumace y, u, w jsou nepovinné, počet operandů v sumacích je omezen na 255. Neuváděním některých sumací a variací počtu operandů v sumacích lze tedy získat z této formule obrovské množství speciálních případů této formule.

Podstatné zjednodušení všech operací na větách je zajištěna standardizací struktury věty. Hlavními komponentami všech vět je :

- znakové pole klíčů, klasifikátorů a indikátorů,
 - 2 nepovinná textová pole
 - nepovinná dvourozměrná matice (v speciálním případě matice o jediném prvku, zápis čísel ve tvaru packed, délka prvků 1 - 8 bytů).
- Celková délka polí ad a), b) je omezena na 255 bytů, celková délka věty je omezena operačním systémem na 32K.

Operace na matici jsou koncipovány jako operace maticové, řádkové, sloupcové a prvkové. Uvedená výpočtová formule může být předepsána jako řádková, sloupcová nebo prvková. Řádkovou resp. sloupcovou operací se rozumí operace, která se opakuje pro všechny nebo některé prvky řádku resp. sloupce, přičemž případ "některé" se zajišťuje prostřednictvím neúplných cyklů.

6. Okolí jádra PARAM II

Nejdůležitějšími a nejsložitějšími prvky soustavy PARAM II jsou kompilátory, z toho jeden společný pro programy jádra a jeden pro tiskové programy. Zdrojové formule (vstup do kompilátorů) a formule v interním tvaru (výstup z kompilátoru) jsou drženy jako soubory. (Pozn.: Operační systém BS 2000 nemá knihovny pro zdrojové programy a hotové programy). Lehké syntaktické chyby lze opravovat interaktivně v průběhu kompilace.

Důležitou složkou soustavy jsou dva generátory tiskových sestav. Variabilita tiskových sestav je umožněna konceptem tiskového standardu, tedy obdobně jako u datových souborů. Tiskové standardy jsou v PARAM II pojaty značně obecně, takže počet tiskových standardů je prakticky neomezen. Uživatelský komfort je zajišťován prostřednictvím knihoven textů a rastrů. Jednou sestavené texty a rastry lze mnohonásobně využívat, lze z nich vytvářet varianty a verze. Tiskové sestavy se vytvářejí řečeno zjednodušeně, vyplněním připravených rastrů.

Další programy slouží pro zavádění dat do systému (ze štítků, terminálu, magnetické pásky), pro vytváření standardních souborů ze štítků nebo pásky, pro převod nestandardních souborů do standardního tvaru a obráceně.

7. Teoretické zázemí PARAM II

Je nutno nejdříve zpřesnit, že programy uváděné v tomto referátu představují pouze úvodní triviálnější část připravovaného komplexu.

Až na program MNOJ (množinové operace) uvedená část PARAM II nepředpokládá od uživatelů žádnou vybraněnou ideologii, t. j. pojetí funkcí těchto programů je akceptovatelné jak pro čisté praktiky, tak pro stoupance některého abstraktního datového modelu (relační, síťový, entitní, entitno-relační). Obtížnost zvládnutí PARAM II pro začátečníky je srovnatelná nebo menší než obtížnost vyššího programovacího jazyka. Zkušenější programátor rychle identifikuje v PARAM II svoji osobní zkušenost a vlastní postupová schemata.

Jinak je to s množinovými operacemi a s dalšími připravovanými programy. Tyto jsou tvořeny v duchu entitno-relačních datových modelů (známé jako

Chenovy E - R modely). Kritickým okamžikem v používání PARAM II je okamžik, kdy uživatel má pochopit, že aktualizace souboru je nutno chápat jako množinové operace.

8. Co může znamenat PARAM II pro praxi

PARAM II má nebo může mít řadu dopadů na projekčně-implentační činnosti :

a) dopad metodický, projektant je veden k entitno-relačnímu pojetí datového modelu,

b) produktivita implementace může být až řádově vyšší ve srovnání s. klasickým programováním,

c) vzhledem k vysoké produktivitě implementace jeví se rozumné slučovat projekci a implementaci,

d) důsledným využíváním superpozice a slučováním funkcí lze docílit mnohonásobné snížení množství operací vstupu / výstupu,

e) je značně omezena možnost psaní špatných programů, což má dopad na: ladění a efektivnost výpočtů.

Zcela úmyslně nemluvíme o výhodách, ale pouze o možných dopadech. Nelze zcela bezpečně tvrdit, zda uvedené dopady jsou předností PARAM II.

9. Možnosti využití

PARAM II je pojat dosti široce a modulárně. Implementace některých modulů resp. funkcí je pouze otázkou času a realizačních kapacit. Už v uvedené sestavě lze pomocí PARAM II realizovat velmi efektivně a bez větších nároků na práci a strojový čas netriviální úlohy sekvenčního zpracování dat. Speciálně lze v PARAM II dobře realizovat úlohy statistického výkaznictví, statistické analýzy, v informačních a řídicích systémech podniků až ústředních orgánů.

Lze připustit, že v PARAM II jsou dílem promítnuty výhradně místní potřeby, osobní vkus a kompromisy autorů koncepce nebo málo domyšlené konvence a koncepce, jenž mohou mít dopad na využívání a doplňování systému.

10. Plány do budoucnosti

Řečeno v terminologii E - R modelů stávající část PARAM II umožňuje manipulace pouze s entitními soubory. Pro komplexní aplikace soustava by měla být doplněna o prostředky pro komplexní manipulace s entitními a vztahovými soubory. Navíc soustava by měla být doplněna o prostředky pro interaktivní

práci s datovými modely.

Za tímto účelem měla by se nejdříve zvládnout metodika budování a využívání datových modelů. Výsledkem tohoto úsilí by měla být práce metodického charakteru pojednávající o konceptuální analýze a syntéze systému pro zpracování dat.

Nakonec připadá do úvahy procedurální trojúrovňový programovací jazyk, jehož nástin už vlastně byl uveden při popisu funkcí programů. V tomto jazyce rozlišovaly by se tři hierarchické úrovně objektů, operací a funkcí. PARAM II je vlastně jednou z možných velmi omezených, zato jednoduchých, implementací tohoto jazyka. Možnosti takového jazyka implementovaného v plném rozsahu by měly pokrývat rutinním způsobem současnou potřebu praxe. Pod "současnou potřebou praxe" můžeme rozumět řešení úloh srovnatelných co do náročnosti například s úlohou o kritické cestě.

11. Závěr

PARAM II byl uveden do používání v srpnu 1981. Jakékoliv dalekosáhlé závěry o využitím přístupu a o vlastnostech a použitelnosti samotného PARAM II jsou předčasné.

Na realizaci projektu se kromě autorů referátu podíleli ještě : Ing. Jan Zapletal, Ing. Jaroslav Zachariáš, Ing. Vladimír Aubrecht, Ing. František Fendrych, Vladimír Jítka (FMF), RNDr. Pavel Drbal, CSc., Ing. Zdeněk Bohuslav, CSc., Ing. Jindřich Šilhán, Ing. Karel Sokol (VÚMS).

12. Literatura

- /1/ Chen, P. P. S. : The entity-relationship model-toward a unified view of data. ACM TODS, Vol. 1, No. 1, March 1975, pp. 9 - 36.
- /2/ Astrahan, M. M., et al. : System R: Relational approach to database management. ACM TODS, Vol. 1, No. 2, June 1976, pp. 98 - 137.
- /3/ Scheber, A., Štunc, J. : Relačný model báz dat. Informačné systémy 5, 1976, č. 3, s. 281 - 294.
- /4/ Štunc, J. : Čo je relačný model? Informačné systémy 8, 1979, č. 3, s. 259-268.
- /5/ Scheber, A., Steiner, J. : Databázové systémy - niektoré aktuálne smery vývoja I. Informačné systémy 9, 1980, č. 1, s. 61 - 84.
- /6/ Scheber, A., Steiner, J. : Databázové systémy - niektoré aktuálne smery vývoja II. Informačné systémy 9, 1980, č. 2, s. 171 - 194.