

Klasifikace systému Redap

Jan Šantl, *Ing. Mat., CSc*

VJSE Běchovice

Úvod

Cílem příspěvku je stručně rekapitulovat přístupy ke zpracování hromadných dat a databázové prostředky, které jsou v ČSSR k dispozici pro počítače SMEP a osmibitové mikropočítače. Tyto prostředky jsou stručně komentovány autorem příspěvku, který je současně stoupencem relačních technologií zpracování hromadných dat, autorem Redapu /implementace relačního modelu na počítačích SMEP/ a programátorem v systému Redap při budování ASŘ mateřské organizace. Z toho vyplývá určitá tendenčnost celého příspěvku, který však, jak doufám, může kritickému čtenáři pomoci k doplnění mozaiky znalostí i v tomto mnohokrát zpracovávaném tématu.

Zpracovatelské technologie

Hromadná data se zpracovávají na počítačích zhruba 20 let. Během této doby se vykrytalizovalo několik technologií zpracování hromadných dat. Každá technologie má své stoupence a odpůrce a to zejména proto, že žádná není absolutně obecně použitelná. Tyto technologie lze rozdělit do dvou základních tříd a to:

- individuální technologie, tj. data v izolovaných souborech a individuální techniky programování a přístupu k datům podporované vyššími programovacími jazyky jako Fortran, Pascal, Cobol, PL 1 ap.
- databankové technologie, tj. data v souborech se speciální strukturou, ovládané speciálními softwareovými prostředky - tzv řízením báze dat. Aplikační programy se však i nadále píšou ve vyšších programovacích jazycích jako u individuálních technologií. Obvykle je do těchto programovacích jazyků vnořena část příkazů řízení báze dat.

Obecně se konstatuje, že individuální technologie jsou vhodné

pro dávkové zpracování s jednoduššími vazbami mezi daty a že databankové technologie jsou vhodné pro interaktivní zpracování se složitými vazbami mezi daty.

První technologií dostaneme relativně rychle a levně první výsledky, ale po relativně krátké době se dostaneme do značných potíží při snaze vícenásobného využití uložených dat a nutnosti restrukturalizace datových souborů.

Druhá /databankové/ technologie je náročná na kvalifikaci personálu a je rovněž náročná i v oblasti návrhu báze dat. Zjistilo se, že obecně všechna data organizace nelze umístit do jedné báze dat, protože by jinak došlo k zahlcení systému řízení báze dat. Obvykle je tedy databanková technologie doprovázená i individuálními technologiemi zpracování hromadných dat.

Dělení bází dat

Databankové technologie se podle způsobu popisu dat a vazeb mezi daty v zásadě dělí na:

- hierarchické /též stromové/
- síťové
- relační

Každý typ báze dat má své prostředky k popisu dat a manipulaci s daty. V hierarchické bázi dat lze modelovat vazby mezi objekty 1:n /každý syn má právě jednoho otce/, v síťové bázi dat lze modelovat vazby mezi objekty m:n /otcové mohou mít více synů, ale i synové mohou mít více otců/. Z toho vyplývá, že hierarchická báze je zvláštní případ síťové báze dat.

Charakteristické pro hierarchické a síťové báze dat je tzv proces návrhu báze dat. Tato činnost je nesmírně náročná a odpovědná, protože na kvalitě návrhu báze dat záleží podstatně úspěch nebo neúspěch přechodu na databankovou technologii. V návrhu báze dat jde stručně řečeno o to, zjistit které informační objekty se v bázi dat budou vyskytovat, jaké budou mít vlastnosti a jaké vazby mezi objekty a na základě jakých

vlastností se budou realizovat. Každé opomenutí vazby či podcenění frekvence uskutečňování informační vazby stojí v budoucnosti mnoho strojového času. Náprava je sice možná restrukturalizací báze dat, ale ta vede obecně k závažným důsledkům pro již vytvořené aplikační programy.

V 70. letech E.F.Codd formuluje teorii tzv relační báze dat, která by měla odstranit nevýhody síťového modelu a to zejména v oblasti návrhu báze dat. Upustíme-li od rigorózní matematické formulace relační báze dat, pak se dá říci, že relační báze dat se skládá z izolovaných datových souborů. Na rozdíl od obyčejného souboru však datový soubor má někde /na začátku nebo v jiném souboru/ uložen svůj popis, tj. jaké položky jsou ve větě souboru uloženy a jakého jsou typu. Vazby mezi větami souborů se uskutečňují na základě shody ve vybrané skupině položek. Tedy aniž bychom cokoliv dopředu projektovali, můžeme pomocí funkcí v relační bázi dat propojit cokoliv s čímkoliv na základě shody ve vybraných vlastnostech informačních objektů reprezentovaných větami souborů.

Přes svou zdánlivou jednoduchost se relační model začíná prosazovat teprve po více jak 10 letech od formulace matematické teorie. Bylo to způsobeno obtížností implementace a následné průchodnosti báze dat /kritizovala se zejména doba odezvy jedné transakce/. Nástup výkonnější výpočetní techniky spolu se snižováním cen hardware však umožnil, že relační báze dat se při zpracování hromadných dat čím dále tím více prosazuje. Hlavní motivací je zvýšení produktivity práce aplikačního programátora.

Databázové systémy v ČSSR pro mini a mikropočítače

Minipočítače SMEP

BTS - relační

DIAMS 1 a 2 - v podstatě relační, 23 instalací

MBBS - relační, 1 instalace

DABASYS - síťový, 0 instalací

GIN - v podstatě relační
REDAP - relační, 94 instalací

8. bitové mikropočítače

DBASE II - relační

/uvedené počty instalací některých systémů jsou převzaty ze studie ÚVVTŘ Praha "Stav využívání systémů řízení báze dat v ČSSR" z listopadu 1986/

System Redap má tedy na počítačích SXEP největší počet instalací ze všech nabízených databázových systémů. Šla uživatelé systémů Redap vyzkoušela i několik dodávaných databázových systémů a o zkušenosti s jejich nasazením se s námi podělili. Následující stručné hodnocení obsahuje zčásti i tuto zobecněnou zkušenost.

- DTS programování snadné, interaktivní přístup podporován. V multiuživatelském prostředí málo průchodný, dva uživatelé s obtížemi, tři zahltili systém. Malá podpora interaktivního programování, časté přeplnění vnitřních bufferů při složitějších úlohách, nízký výkon v dávce.
- DIAMS 1 a 2 programování snadné, interaktivní přístup podporován. Problematičnost nasazení spočívá ve skutečnosti, že se jedná o samostatný operační systém neslučitelný s běžně užívaným DOS RV. Během dne je pak nutno provádět komutace mezi zpracováním v DIAMSu a v DOSu. Multiterminálový provoz tím přirozeně trpí. DIAMS je výhodný pro interaktivní přístup k bázi dat, ale nevýhodný pro dávkové zpracování díky značně atomizované bázi dat na všech dostupných diskových médiích. Počet přístupů na disk je pak v dávce o jeden až dva řády větší než by bylo nezbytné nutné.
- MBBS programování pracné, mírná podpora interaktivního přístupu k bázi dat. Programy se píšou v běžných programovacích jazycích s využitím volání funkcí řízení báze dat. Založen na komprimované bázi dat, připouští však maximálně 256 souborů v bázi dat. O průchodnosti systému v dávce nemám informace.

- DABASYS programování pracné, podpora interaktivního přístupu k bázi dat chybí. Vyžaduje 1 Mbyte operační paměti. Aplikační programy se píšou v Cobolu. Vychází ze stejné normy jako IDMS na počítačích JSEP.
- GIN s tímto systémem nemám žádné osobní zkušenosti ani zprávy od našich uživatelů. Z materiálů, které mám k dispozici vyplývá asi toto: Definují se položky, věty, soubory a formuláře. Všechny prvky se označí číselnými identifikátory. Součástí systému jsou parametrické programy pro instalaci a kopírování bází dat, výpis stavu databázových souborů a prvků pro správce báze dat, pro vstup a aktualizaci dat, třídění, zpracování a tisk sestav. Existují též programy pro čtení nebo vytváření magnetických pásek pro systém JSEP. Složitější úlohy se programují v Asembleru, jednodušší ve Fortranu při využití knihovny řízení báze dat systému GIN. Současně lze využívat 4 diskové jednotky, soubory se definují předem jako prázdné a neměly by mít více jak 32767 vět. Deklarován jako vhodný systém pro řízení v reálném čase.
- DBASE II poměrně komfortní implementace relačního modelu na mikropočítačích. Programování snadné, interaktivní přístup podporován. Hlavní nevýhodou je ukládání číselných informací v textové formě a tedy neúměrně velká spotřeba omezené disketové paměti. Další nevýhodou je malá výkonnost systému v dávce. Schází možnost virtualizace tabulek a tedy nutnost dalšího místa pro přípravu tisku.

REDAP

Je to poměrně netypická implementace relačního modelu. Systém je koncipován jako nástroj aplikačního programátora. Původním cílem byla podpora dávkového zpracování hromadných dat. Redap vcelku přirozeným způsobem zobecňuje souborový přístup k zpracování hromadných dat /individuální technologie - viz začátek příspěvku/ a povyšuje ho k téměř k databázovému zpracování prakticky bez zvýšení kvalifikačních nároků na aplikačního programátora.

Odolnost programů v systému Redap vůči změnám v bázi dat podporuje budování ASŘ od zdola nahoru bez počátečních nároků na náročné a nákladné analýzy toku dat. ASŘ se přirozeně rozvíjí v souladu s rozvíjejícím se chápáním problematiky jak programátora, tak uživatele. Tato okolnost je ještě umocněna několikanásobným zvýšením produktivity práce aplikačního programátora oproti použití individuálních technologií.

Redap je výkonný v dávce /připouští teoreticky 10^6 vět/ a poskytuje i rozumné doby odezvy při interaktivním přístupu až do oblasti 10000-20000 vět v jednom souboru /hledání 15 až 30 sekund/. Je úsporný při ukládání dat i programů. Umožňuje interaktivní programování vhodné obzvláště při řešení ad hoc dotazů. Zkušenost ukazuje, že odborní referenti jsou schopni v krátké době převzít do své správy celý provoz konkrétního subsystému.

Programy v systému Redap jsou stručné a mají autodokumentační charakter. Z toho vyplývá i poměrně jednoduchá dokumentace /pokud je nutno vůbec nějakou dělat/. Z toho vyplývají určité rozpaky uživatelů systému Redap. Prakticky všude chybějí projekty subsystémů a dokumentace k hotovým programům. Všeobecně se konstatuje, že projekty a dokumentace je daleko pracnější než fyzické vytvoření programů. Na rozdíl od běžných přístupů máme fungující subsystémy bez projektů a dokumentace, zatím co jinde často vidíme dokonalé projekty bez programů.

v. Jako autor Redapu vidím dnes hlavní nevýhodu Redapu jednak nemožnost aktualizace jednoho souboru z více míst /více-násobný přístup na zápis/ a jednak v tom, že relační pohled je někdy pro uživatele nepřirozený. Bylo by zapotřebí na vstupu i výstupu připustit v některých případech hierarchické vazby mezi větami různých souborů. Oba zmíněné problémy však neodporují celkové koncepci systému a budou postupně v dalších verzích řešeny.

V soutěži s jinými dodávanými systémy na počítačích SMEP vítězí Redap podle mého názoru proto, že respektuje důsledně československé poměry a to zejména:

- omezené diskové kapacity /úsporné ukládání dat i programů/
- pozdlý hardware /pečlivé zpracování hlavních cyklů/
- malá operační paměť /vhodné segmentace služebních programů/
- dobré personální obsazení VS /orientace systému na aplikačního programátora a ne jak je u relačních bází dat zvykem orientovat se na uživatele neprogramátora/

K tomu je zapotřebí připočíst i další vlastnosti jako:

- nízké procento selhání systému při řešení problematiky zpracování hromadných dat tj. nutnost použít jiné nástroje.
- vysoký výkon v dávce
- rozumná doba odezvy při interaktivním přístupu
- vysoká produktivita práce aplikačního programátora
- snadnost ovládání hotových produktů pro uživatele neprogramátora
- velká odolnost programů vůči změnám v bázi dat
- kvalitní servis /školení, konzultace, stáže/
- dobrá dokumentace udržovaná v aktuálním stavu
- rychlý rozvoj systému reagující na problémy uživatelů

Je jasné, že řadu deklarovaných předností Redapu poskytují i jiné systémy často na lepší úrovni. Ovšem o úspěšnosti nebo neúspěšnosti toho kterého systému rozhoduje celková vyváženost z hlediska výhod a nevýhod.

MikroRedap

Ve druhém čtvrtletí 1986 byl zahájen převod systému Redap na 8. bitové počítače na bázi procesoru I8080 a mající systém Mikros /nebo CP/M/. V okamžiku psaní tohoto příspěvku je převod prakticky ukončen. Řešitelskému kolektivu se podařilo zajistit prakticky 100% kompatibilitu dat i programů v systému Redap mezi minipočítači SMEP a 8. bitovými mikropočítači. Tento výsledek je pozoruhodný i z toho hlediska, že zatím u nás neexistuje žádný databázový systém fungující na odlišných tří-

dách počítačů.

Dnes tedy to co je hotové pod systémem Redap na minipočítačích SMEP bude fungovat i na mikropočítačích a programátor z počítače SMEP znalý systému Redap vystačí se svými znalostmi i na obsluhu 8. bitových mikropočítačů. Přitom Mikro-Redap je jako Redap na SMEFu výkonný a současně nenáročný na diskové či spíše disketové kapacity.