

Najvyšší rychlostný stupeň firemných analýz

» Svet technológií sa stále viac točí okolo magickkej formuly in-memory. Tá je vďaka presunu operácií z pevných diskov do operačnej pamäte etalónom rýchlosti a efektivity. Azda najvýraznejšie sa prínos in-memory nástroja prejavuje v oblasti business intelligence a firemných analýz, kde zrýchľuje a skvalitňuje interpretáciu dát. S kvalitným in-memory môžete nájsť netušené možnosti na ďalší rozvoj, ako aj nové zdroje príjmov.

Teória je jasná, ale s praxou môžu prísť prvé problémy. Čo ak máte databázu s veľkosťou 100 GB a serverovú RAM so 40 GB? Alebo dokonca len notebook či desktop, kde objem operačnej pamäte zriedka prekročí 16 GB. Tradičná cesta navrhuje „ťahat“ dáta z pevného disku, na výsledky si však počkáte.

Pevné disky vs. in-memory

Tradičné databázové systémy fungujú na princípe dát uložených na pevných diskoch, pričom používateľ tieto údaje dopytuje. Ak chce vidieť napríklad predaj produktov podľa zákazníkov, databáza túto informáciu o nejaký čas poskytne. Ale práve čas je pre biznis kritický, odpoveď o hodinu vám nepomôže, ak ju potrebujete ihneď. Nehovoriac o tom, že časté otázky tohto druhu vás môžu výrazne zdržovať pri práci.

Naproti tomu súčasné pokročilé in-memory technológie dokážu výrazne znížiť ich veľkosť, vďaka čomu môžete s nimi pracovať v operačnej pamäti. Pri tých najlepších sa môžete dostať na desatinu pôvodnej veľkosti dát či dokonca ešte menej. Aj po nástupe diskov SSD platí, že operačná pamäť je o krok vpred. Argument, že RAM má oproti diskom výrazne nižšiu kapacitu, už neobstoí, čo je vizitka účinnej komprimácie a optimalizácie dát in-memory technológiami. Práve to je najväčšia konkurenčná výhoda in-memory nástrojov a tí, ktorí sa rozhodli ju na začiatku rozvíjať, sú dnes lídrami s veľkým trhovým náskokom.

Rekordné štyri terabajty dát

Pri tradičných databázových nástrojoch môžu v závislosti od náročnosti analýzy a veľkosti databázy reakcie systému trvať minúty až hodiny, kvalitné in-memory však tie isté operácie zvládne v rozsahu milisekúnd až sekúnd, pritom počet záznamov nie je zásadné úzke miesto. Navyše bez toho, aby museli byť tieto dáta spracúvané na „sci-fi“ serveroch. Pokojne to môžu byť aj stovky miliónov či dokonca miliardy a stále je reč o sekundových reakciách.

Ako príklad môžeme uviesť spracovanie dátového setu s veľkosťou štyroch terabajtov technológiami visual analytics (známe aj ako data

discovery) v rámci in-memory. Podarilo sa to na serveri s operačnou pamäťou 512 GB RAM. Štvorterabajtová databáza bola skomprimovaná tak, aby sa zmestila do RAM-ky, a aj v tomto prípade platilo, že reakcia bola rádovo niekoľko sekúnd. Keby sa táto analýza mala realizovať tradičným spôsobom na pevných diskoch, pri tejto veľkosti databázy by sa čakacia doba na výsledok mohla počítať aj v hodinách.

Indexovanie dát

In-memory technológia je z technického hľadiska založená na indexovaní údajov. Samozrejme, dôležitá je aj metodika vytvorenia dátového modelu. Keďže pri veľkých dátach ide o pomerne komplikovanú záležitosť, vyžaduje si naozaj odborné podkutého, certifikovaného špecialistu. Cieľom je dáta pripraviť čo najlepšie, pričom vo všeobecnosti platí, že komprimácia bude tým efektívnejšia, čím lepšie sa údaje pripravujú. Kľúčové je pochopiť, čo sa v tabuľkách nachádza, čo sú kľúčové informácie, ako ich prepojiť a ako bude dátový model vo finále vyzerať. Najlepšie z in-memory nástrojov indexujú každý jeden údaj v bitovej sústave, teda v jednotkách a nulách a vo vedľajšej tabuľke si ukladajú celú hodnotu. Takýmto spôsobom in-memory optimalizuje všetky pripojené dáta.

Treba podotknúť, že dátový model sa dá urobiť dobre, ale aj pokaziť. Veľakrát sme prerábali riešenia, ktorých problémom boli nedostatočne optimalizované dáta. Ak bol napríklad dátový model urobený zle, 100 MB na pevnom disku sa v RAM rozvinulo napríklad do 2 GB. To spôsobilo, že dátový model zbytočne zahltal pamäť a všetky aplikácie bežali pomaly, čas reakcie a výpočtov sa zdvojnásobil. Všetko teda stojí na tom, ako dobre rozumieme údajom – či správne chápeme, ktoré polia sú číselné, ktoré sú kľúčové, ktoré textové, aká je tam kardinalita a podobne.

Zo 100 % na 6 %

Na ilustráciu toho, kam sa môžete dostať kvalitnou optimalizáciou a komprimáciou dát, si načrtnime príklad z našej praxe. Išlo o nasadenie nástroja visual analytics s integrovanou in-memory funkcionalitou vo vedúcej energetickej spoločnosti.

Tá disponovala dátovým setom vo veľkosti asi 1,2 miliardy údajov a vďaka inovatívnym technológiám sa nám ich objem podarilo skomprimovať na pätinu pôvodnej veľkosti. Po následnej logickej agregácii dát výsledná komprimácia dosiahla úroveň 6 %. In-memory teda nie je len o komprimácii, tento krok predchádza ešte „čistenie“ dát, aby ostali len tie relevantné. Optimalizácia dát je veľmi dôležitá

časť celého procesu, ktorá pomôže výkonu a urýchli spracovanie dát. Na ich selekciu je však potrebný ostrieľaný partner, zdatný v business consultingu. My vychádzame z očakávaní a potrieb zákazníka, ako aj z nášho know-how, čiže vieme, ktoré informácie sú podstatné a ktoré nie.

Ohľadom optimalizácie dát ešte jedna poznámka o tvorbe dátového modelu. Ak máte k dispozícii kvalitný in-memory nástroj, môžete začať validným minimom alebo optimom a ďalšie relevantné dáta môžete pripojiť neskôr podľa aktuálnych potrieb. Naopak, pri slabších nástrojoch od začiatku musíte ísť maximalisticky, pretože je zložité niečo do nich dodatočne pridávať.

Na in-memory platforme záleží

In-memory technológie najčastejšie fungujú v rámci serverových riešení. Výhodné je, že v takomto prostredí sa na údaje môže pripojiť a pracovať s nimi naraz väčšie množstvo ľudí a nie je problém so zachovaním jednej verzie pravdy.

Jedným dychom však treba dodať, že niektoré vyspelejšie in-memory nástroje sú schopné fungovať aj na bežnom desktope alebo notebooku, nie sú totiž viazané na serverové prostredie. Aj keď to môže znieť prekvapujúco, s kvalitnými in-memory nástrojmi môžete pracovať s miliardami záznamov alebo gigabajtmi dát pri „bežných“ HW nárokoch, rádovo v jednotkách GB RAM. Jeden maloobchodný reťazec takto na bežnom kancelárskom laptope analyzoval miliardový dátový set. Po komprimácii dosiahol objem 13 GB, postačilo tak 16 GB RAM.

Vyberajte správne

V rámci BI a vizuálnych analýz je kľúčová rýchla a flexibilná práca s údajmi, aj preto je dnes in-memory v trende. Treba však povedať, že hoci sa in-memory technológiami chváli viacero vendorov, nie každý prinesie sľubované výsledky. Kým niektorí vyvíjajú in-memory technológie už dve desiatky rokov, iní začínajú len teraz a rozdiely vidieť na prvý pohľad – napríklad v schopnosti komprimovať databázy, ukazovať dáta zmysluplne a v kontexte biznisu či okamžitej zrozumiteľnosti. Pri výbere analytického softvéru sa oplatí preveriť úroveň in-memory technológií a to, akým spôsobom ukladajú údaje.

Pri výbere riešenia sa treba pýtať aj na to, ako dlho konkrétna značka svoju in-memory technológiu vyvíja. Firmy, ktoré sa tomu venujú dvadsať rokov, sú technologicky oveľa ďalej ako tie, ktoré sa na tejto vlně vezú len tri či päť rokov. Nič neukáže lepšie schopnosti nástroja ako test na reálnych dátach. Otestujte ho nad niektorou zo svojich databáz a správny nástroj v priebehu hodín až dní prinesie zásadné výsledky.

» MARTIN KOSTIČ,
riaditeľ EMARK, s. r. o.

