

Test výkonnosti při indexování bitovou mapou

Caché versus Oracle

9. srpna 2002

Indexování bitovými mapami je sice už dlouho využíváno pro zrychlení vyhledávání v systémech datových skladů, avšak nedostatečná výkonnost při operacích aktualizace bránila jeho využití v aplikacích s transakčním zpracováním. Jako reakci na rostoucí potřebu analýz v reálném čase, pro kterou je třeba používat složité dotazy pracující s živými daty, přináší systém Caché 5 novou technologii transakčního indexování bitovou mapou, která se vyznačuje radikálně zvýšenou výkonností při aktualizacích. Tato jedinečná inovace umožňuje použití indexování bitovou mapou i v aplikacích, kde hraje výkonnost při transakčním zpracování důležitou roli.

Společnost InterSystems demonstrovala ohromné zvýšení výkonnosti systému Caché 5 při indexování bitovou mapou v transakčním zpracování na níže popsanych testech.

Definice databáze

Při testech bylo použito jedné tabulky o 9 sloupcích definované takto:

```
Create table Osoby(  
    ID INTEGER Primary Key,  
    Jméno VARCHAR(50),  
    Titul VARCHAR(50),  
    Pohlaví VARCHAR(1),  
    Stav VARCHAR(2),  
    Plat INTEGER,  
    Věk INTEGER,  
    Kód VARCHAR(10),  
    BarvaVlasů VARCHAR(10)  
)
```

Bitmapové indexy byly vytvořeny pro 7 sloupců (pro všechny kromě ID, jehož hodnoty jsou jedinečné, a Jméno, jehož hodnoty jsou skoro jedinečné. (Bitmapové indexy se obecně nedoporučují pro sloupce, které obsahují velké množství jedinečných hodnot.) Indexy byly definovány příkazy SQL s touto syntaxí:

```
create bitmap index ITitul on Osoby(Titul)
```

Množství dat v databázi

Do databáze bylo nahráno 10 000 000 řádků. Jak ukazuje následující tabulka, množství různých hodnot, ze kterých lze vybírat, se v jednotlivých sloupcích značně liší (srovnejte např. sloupce Jméno a Pohlaví).

Sloupec	Počet různých hodnot
Jméno	711 253
Titul	136
Pohlaví	2
Stav	50
Plat	99
Věk	98
Kód	6
BarvaVlasů	5

Testovací prostředí

Testy byly provedeny na systému této konfigurace:

CPU	Duální procesor Intel Pentium 3 Xeon s taktovací frekvencí 2.0 GHz a s 512 kB cache
Paměť	1 GB paměti RDRAM
Disk	Duální harddisk Seagate Cheetah SCSI o velikosti 18,4 GB
Operační systém	Windows 2000 Server (SP 2, sestavení 2195)
Databáze	Caché 5 Oracle 9i Enterprise Edition

Výkonnost při aktualizaci

Výkonnost při aktualizaci byla testována pomocí jednoduchého příkazu UPDATE jazyka SQL:

```
Update Osoby Set Kód = 'A' where ID= :id
```

Tento příkaz byl vykonán (v uložené proceduře, aby byly eliminovány latentní problémy v komunikaci mezi klientem a serverem) 1 000krát a byly přitom modifikovány řádky rozhozené po celé tabulce s 10 000 000 řádků.

Jak ukazují následující výsledky, databáze Caché provádí aktualizace přibližně 300krát rychleji, než databáze Oracle.

	Doba aktualizace 1 000 řádků	Počet aktualizovaných řádků za sekundu
Caché	0,11 s	8 787
Oracle	34,01 s	29

Abychom ověřili, že tento rozdíl byl dosažen právě díky indexování bitovou mapou, provedli jsme test aktualizace také na tabulkách Caché a Oracle bez indexů. Na základě poměru výsledků dosažených s bitmapovým indexem a bez indexu jsme pak spočítali časovou „ztrátu“ způsobenou indexováním bitovou mapou.

	Časová ztráta při indexování bitovou mapou
Caché	1,45
Oracle	6800

Jinými slovy přidání bitmapového indexu k databázi Caché prodloužilo dobu aktualizace o něco méně než o 50%. To samé prodloužilo dobu aktualizace v databázi Oracle **6800krát**.

Výkonnost při vyhledávání

Abychom se ujistili, že výkonnost databáze Caché při aktualizaci není na úkor výkonnosti při vyhledávání, provedli jsme také sérii testů vyhledávání. Obecně řečeno byla výkonnost vyhledávání srovnatelná (v 5 případech byla lepší databáze Caché, v jednom případě databáze Oracle).

Dotaz	Počet vrácených řádků	Doba operace (sekundy)	
		Caché	Oracle

Select count(*) from Osoby where Stav = 'ŽV'	200 108	0,0028	0,0050
Select count(*) from Osoby where Pohlaví = 'M'	4 998 018	0,0056	0,0100
Select count(*) from Osoby where Věk < 40	3 038 594	0,0367	0,0400
Select count(*) from Osoby where Věk between 10 and 40	3 184 366	0,0366	0,0300
Select Jméno from Osoby where Stav = 'ŽV' and Věk = 50	4 448	13,6239	14,0200
Select Jméno from Osoby where (Věk between 5 and 6 or Stav = 'ŽV') and BarvaVlasů='zrzavá'	42 562	17,8577	18,0100