

Analytické dotazy a materializované pohledy

Ondřej Benkovský, Tom Bartoň, Andrea Navrátilová a Vojtěch Frnoch

Popis problematiky

- Analytické dotazy = provádějí komplexní analýzu dat, využití při data miningu
- Využívají často agregační, seskupovací funkce, klouzavé okna, ...
- Provádění takových dotazů může být výpočetně náročné nad neupravenými daty
- OLAP umožňuje uspořádat data v různých dimenzích, tak aby určité analytické dotazy byly efektivnější

Cíle projektu

- Představit podporu materializovaných pohledů v PostgreSQL
- Představit podporu pro OLAP v PostgreSQL
- Implementovat analytické dotazy nad testovacími daty
- Porovnat výkonnost různých implementací analytických dotazů

Představení testovacích dat

- Relace **conn_log** a **service_log**
- Relace **conn_log**, obsahující data o vytvořených spojeních mezi PDA a serverem
- Každé nové spojení vytvoří nový záznam
- Schéma:
conn_log(log_key, sim_imsi, time, car_key, pda_imei, gsmnet_id, method, program_ver)

Představení testovacích dat

- Relace **service_log**, obsahující periodická hlášení aplikací
- Aplikace podá hlášení cca každých 10 minut
- Každé nové hlášení vytvoří nový záznam
- Schéma:
service_log(service_key, car_key, time, app_run_time, pda_run_time, device)

Příklady analytických dotazů

- Pro každou verzi programu zjistit počty různých zařízení
- Pro každou verzi programu zjistit počet restartů jeho programu
- Pro každé zařízení zjistit počet restartů
- Pro každé zařízení zjistit, kdy bylo uvedeno do provozu
- Pro každé auto zjistit počet různých zařízení
- Kolik bylo vytvořeno spojení v různých časových dimenzích (rok, měsíc, den)
- Kolik unikátních zařízení se připojilo v určitých časových dimenzích
- Pro každé zařízení zjistit, kolik hodin bylo používáno

Návrh výkonnostních testů

- Analytické dotazy se implementují nad původními neupravenými daty
- Analytické dotazy se implementují nad původními daty s indexy
- Analytické dotazy se implementují nad upravenými předagregovanými daty
- Porovnání výkonu těchto variant pomocí *EXPLAIN ANALYSE*

Plán projektu

Task Name	Start Date	End Date	Duration	Assigned To
Nastudování problematiky	03/22/17	03/29/17	20h	Všichni
Analýza testovacích dat	03/27/17	03/29/17	10h	Ondřej
Návrh analytických dotazů	03/30/17	04/05/17	10h	Andrea
Implementace dotazů	04/06/17	04/14/17	15h	Andrea
Implementace dotazů	04/06/17	04/14/17	10h	Tom
Implementace dotazů	04/06/17	04/14/17	6h	Ondřej
Validace dotazů	04/15/17	04/19/17	10h	Tom
Testování výkonnosti dotazů	04/20/17	04/24/17	8h	Vojtěch
Vyhodnocení výkonnosti	04/25/17	04/28/17	10h	Vojtěch
Příprava finálního výstupu	04/29/17	05/04/17	10h	Ondřej

