

PODPORA VELKÝCH DAT PRO PIVOT TABULKU

Datum: 19. května 2022

OBSAH

1.1	CO JE NOVÉ (REMOTEOPERATIONS) ŘEŠENÍ A JAK TO FUNGUJE.....	3
1.2	ZÁKLADNÍ POŽADAVKY PRO PROVOZ PIVOT TABULKY V REMOTE REŽIMU.	3
1.3	NASTAVENÍ PIVOTA PRO PROVOZ V REMOTE REŽIMU.	3
1.4	AKTUALIZACE SNAPSHOT TABULEK.....	5
1.5	ZÁKLADNÍ POŽADAVKY PRO SNAPSHOT TABULKY V REŽIMU VIEW.	5
1.6	NASTAVENÍ SNAPSHOT TABULKY JAKO VIEW	5

Aktuální fungování pivot tabulky je takové, že se všechny zdrojové řádky načtou do prohlížeče, kde je Javascript zpracovává do pivot tabulky. Nevýhod je hned několik

- Javascript je jednovláknový a jeho výkon je odvislý od zařízení kde běží.
- Maximální velikost jednoho okna prohlížeče v paměti je cca 2 GB, a tím je limitováno i množství zpracovávaných dat. Aktuálně to je něco okolo 1 mil zdrojových řádků.

1.1 CO JE NOVÉ (REMOTEOPERATIONS) ŘEŠENÍ A JAK TO FUNGUJE.

Remote operations znamená režim, kdy veškeré agregace požadované pivot tabulkou provádí SQL server a na klienta (do prohlížeče) se přenáší už jen výsledek. To umožňuje na SQL serveru zpracovávat i desítky milionů řádků a do prohlížeče přenést už jen stovky nebo tisíce řádků jako výsledek. Rozdíl je obrovský jak v rychlosti, tak i ve velikosti přenášených dat a pak i samozřejmě v rychlosti práci prohlížeče, kde otevřené okno nemá 1,5GB v paměti ale třeba jen 80MB. Při rozpadu v pivotu (drill down) se z SQL serveru načtou už jen příslušné řádky potřebné pro rozpad.

1.2 ZÁKLADNÍ POŽADAVKY PRO PROVOZ PIVOT TABULKY V REMOTE REŽIMU.

- Tabulky jsou umístěné v **PostgreSQL** serveru nebo v **Oracle**
- SQL server je správně nakonfigurován a má dostatečné množství paměti k dispozici. Tady více než jinde platí, že pro práci s velkými daty musí být dostatečně výkonný HW a správně vše nakonfigurováno.
- Jednotlivé sloupce jsou indexovány, což umožňuje rychlé vyhledávání a agregace.

POZNÁMKA

Aktuální rekord je provozování pivot tabulky nad snapshot tabulkou s 22 mil řádků ve 35 sloupcích, přičemž načtení tohoto pivota trvá cca 6 sec.

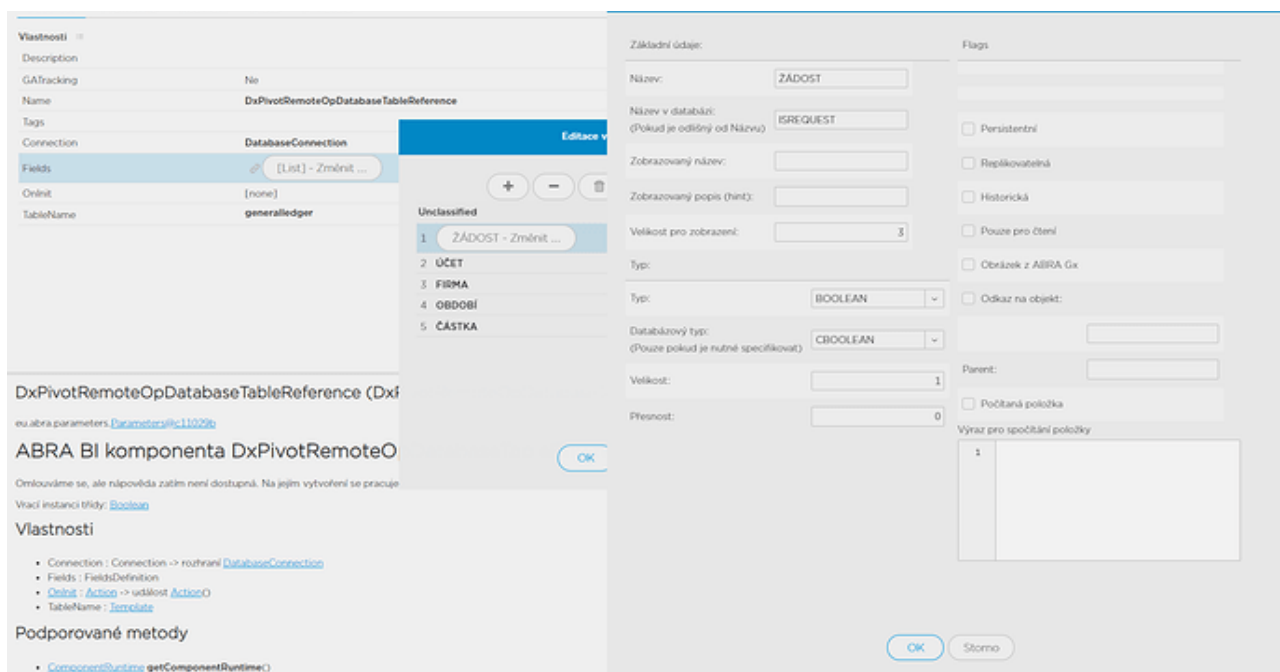
Je třeba mít i dostatečnou kapacitu disku. Snapshot tabulka se 22 mil řádky v 35 sloupcích s indexy nad všemi sloupci zabírá skoro 30GB na disku.

1.3 NASTAVENÍ PIVOTA PRO PROVOZ V REMOTE REŽIMU.

Nastavení se provede zapnutím volby "RemoteOperations" na komponentě DxPivot a vyplněním vlastnosti RemoteOperationsTable – tabulka, nad kterou pivot bude pracovat.

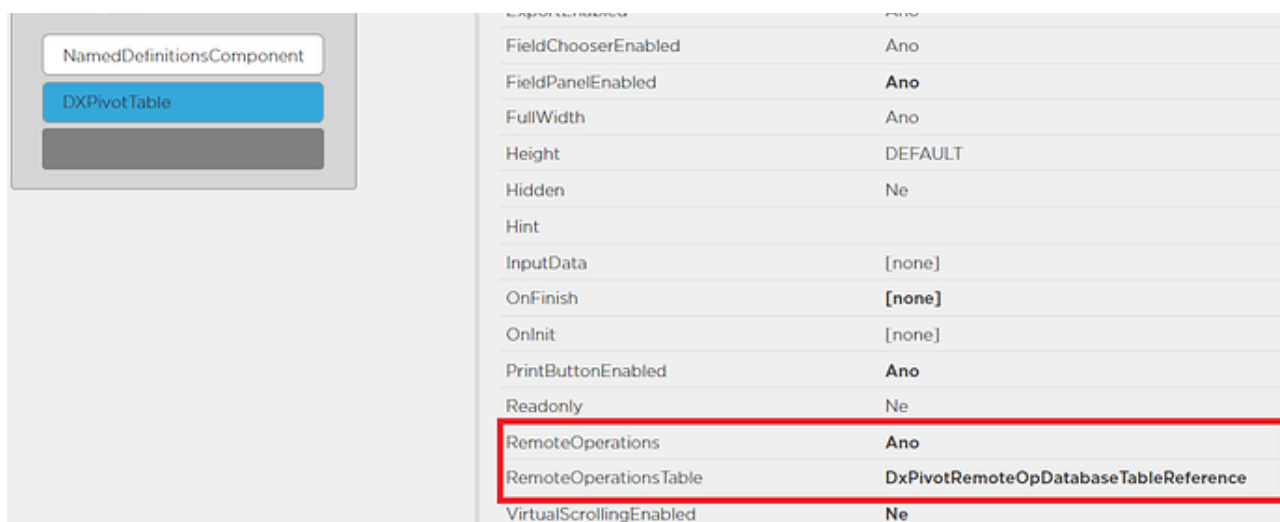
RemoteOperationsTable má dvě implementace:

- **DxPivotRemoteOpSnapshotTableReference** - jednoduchý provider, který pracuje nad snapshot tabulkami a zde se pouze vybere snapshot tabulka.
- **DxPivotRemoteOpDatabaseTableReference** - provider, který zprostředkovává tabulky z jiné databáze. zde je potřeba zadat
 1. Connection - databázové spojení
 2. Fileds - zde se ručně nakonfigurují sloupce, které ze zdrojové tabulky budou nabízet v pivotu
 3. TableName - jméno zdrojové tabulky nebo materializovaného View



1749×946 110 KB

Dále je vše standardní. Pivot si načte strukturu a nabídne sloupce do pivotu. Doporučuje se při výběru sloupců (ve FieldChooser dialogu) vypnout volbu Aplikovat změny hned, jinak se pivot překresluje ihned při přetažení sloupců, a to znamená, že se začínají provádět operace v SQL serveru ihned. A to by bylo zbytečné. Stačí to provést až při ukončení dialogu, když máme vše vybráno.



1026×442 25.9 KB

POZOR

Možná bude potřeba při přechodu na nový režim pivotu znovu nastavit uložené pohledy v pivotu.

1.4 AKTUALIZACE SNAPSHOT TABULEK

S podporou takto velkých dat samozřejmě přichází i problém nahrávání těchto dat do snapshot tabulek. Vždy samozřejmě záleží na složitosti SQL které data do snapshot tabulky načítá. Např. zmiňovaná tabulka s 22 mil řádků nemá až tak složité SQL a je to provozováno proti Oraclu, takže její aktualizace zabere cca 1-2 hod. Jedná se o data několik let nazpět.

Z tohoto důvodu se spolu s novým režimem pivotu objevuje i nový režim snapshot tabulek. Jedná se o snapshot tabulku typu View (Materialized view).

Základní myšlenka celého tohoto systému je v tom, že si tato velká data rozdělím do více snapshot tabulek, které se aktualizují v různých intervalech.

- První největší tabulka představuje data, která jsou napočtena vždy za celý uzavřený rok. To znamená že tato tabulka obsahující např. 20mil řádků se přepočítává jen jednou ročně, a to vždy na začátku nového roku.
- Druhá tabulka obsahuje data za aktuální rok, ale jen za uzavřené měsíce (např. 1.5 mil řádků), takže se počítá jen jednou měsíčně vždy na začátku nového měsíce.
- Třetí tabulka obsahuje data aktuálního měsíce (např. 0,5 mil řádků) a počítá se každý den.

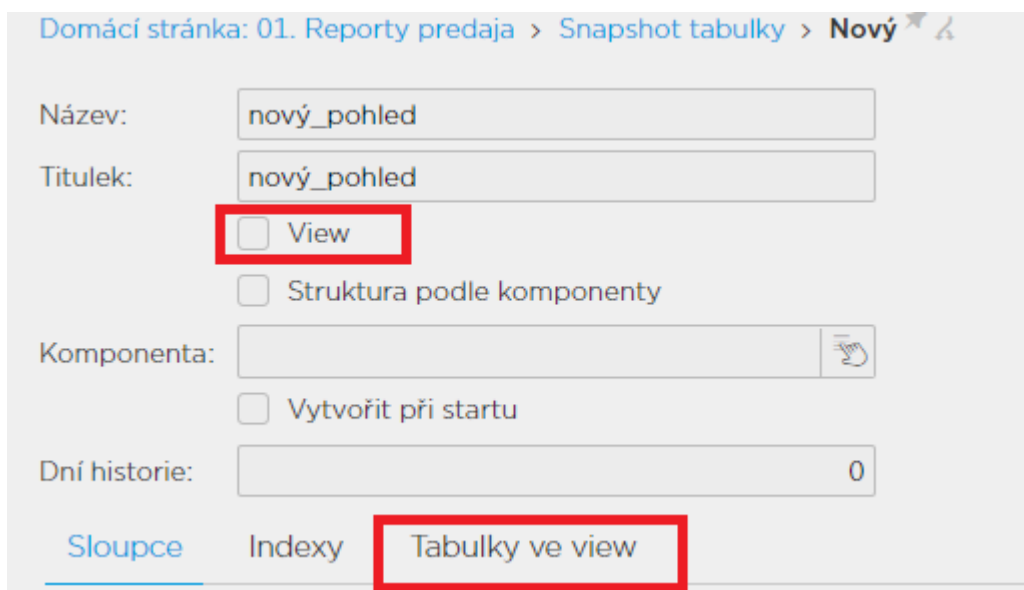
Při změně kterékoliv z těchto tabulek je vyvolána i aktualizace view.

1.5 ZÁKLADNÍ POŽADAVKY PRO SNAPSHOT TABULKY V REŽIMU VIEW.

- Snapshot tabulky jsou umístěné v **PostgreSQL** serveru

1.6 NASTAVENÍ SNAPSHOT TABULKY JAKO VIEW

Zaškrtnutím volby View zapneme na snapshot tabulce tento režim.




Domácí stránka: 01. Reporty predaja > Snapshot tabulky > Nový

Název:

Titulek:

View

Struktura podle komponenty

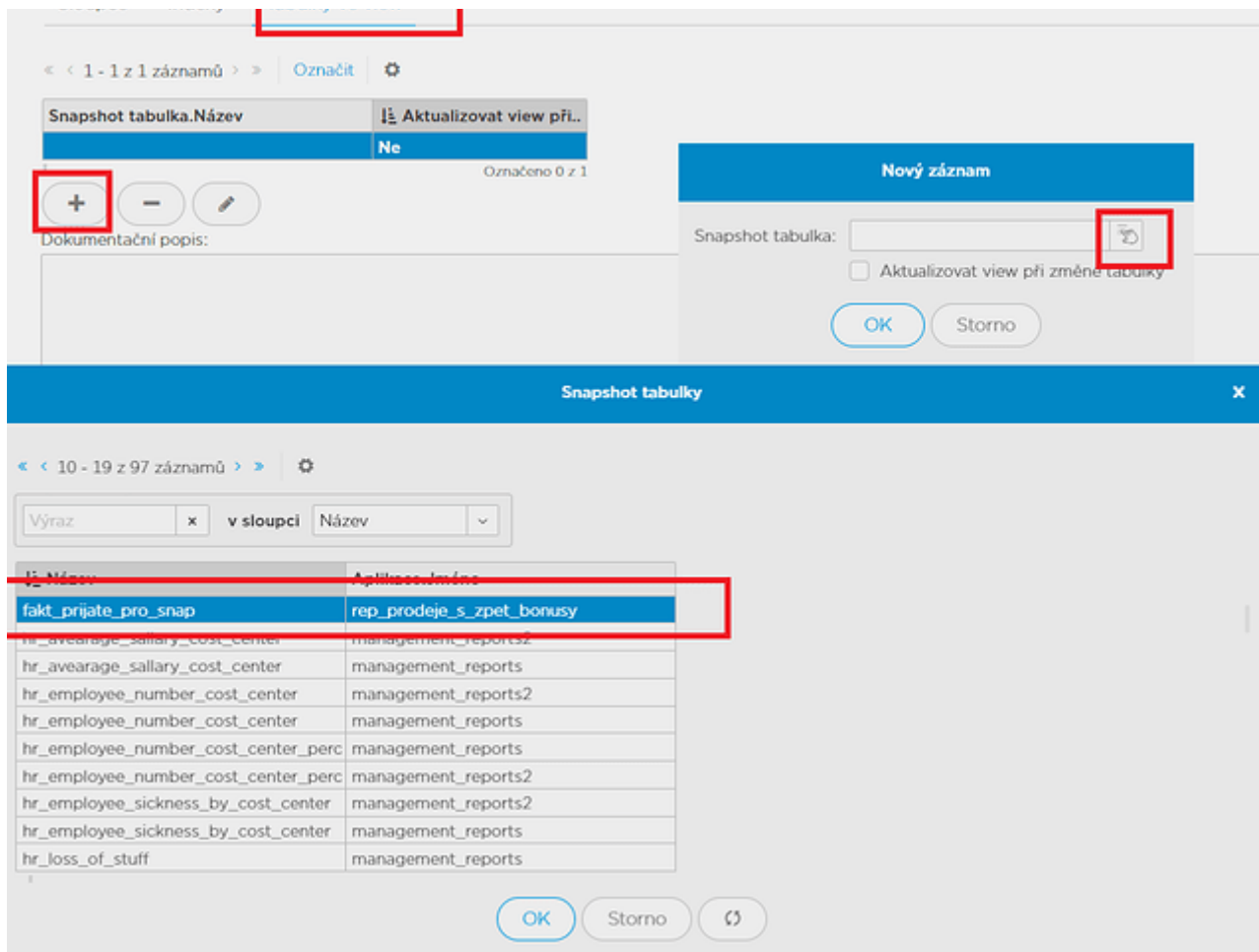
Komponenta: 

Vytvořit při startu

Dni historie:

Sloupce Indexy **Tabulky ve view**

Pak lze přidávat / odebírat jednotlivé tabulky ve view. Zaškrtnutím volby Aktualizovat view při změně tabulky dojde k automatické aktualizaci tohoto pohledu při přepočtení této tabulky. Nemusejí se vytvářet žádné závislé plánované úlohy. Jakmile se změní tabulka odkazovaná ve view, aktualizuje se i celé view.



Snapshot tabulky

Nový záznam

Snapshot tabulka: Aktualizovat view při změně tabulky

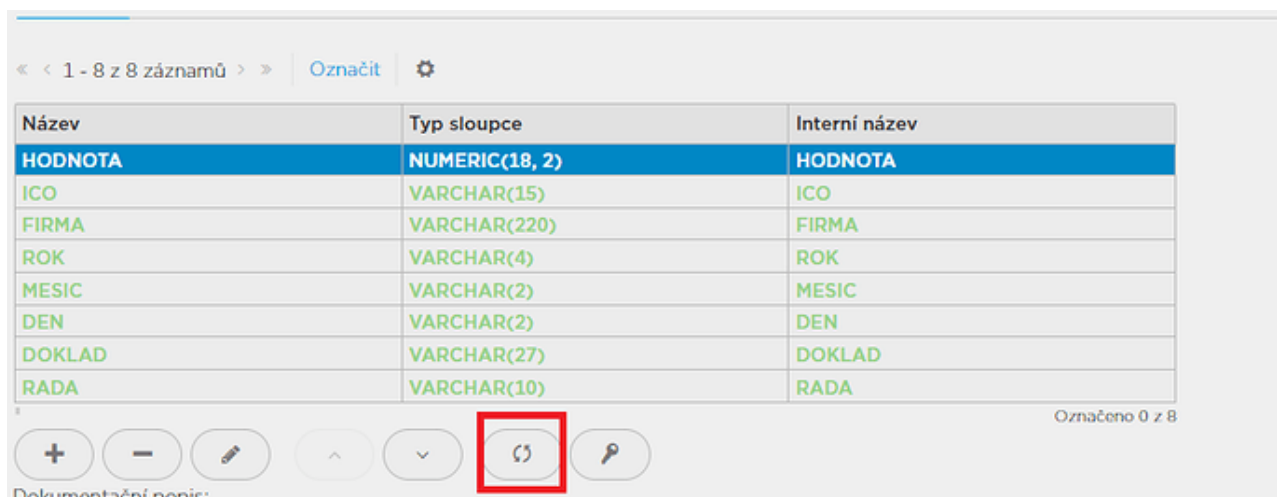
OK Storno

Název	Aplicace/účet
fakt_prijate_pro_snap	rep_prodeje_s_zpet_bonusy
hr_average_salary_cost_center	management_reports2
hr_avearage_sallary_cost_center	management_reports
hr_employee_number_cost_center	management_reports2
hr_employee_number_cost_center	management_reports
hr_employee_number_cost_center_perc	management_reports
hr_employee_number_cost_center_perc	management_reports2
hr_employee_sickness_by_cost_center	management_reports2
hr_employee_sickness_by_cost_center	management_reports
hr_loss_of_stuff	management_reports

OK Storno ↻

1044x814 58.5 KB

Jakmile je ve view zařazena alespoň jedna tabulka, pak lze načíst strukturu podle této první tabulky



1 - 8 z 8 záznamů

Název	Typ sloupce	Interní název
HODNOTA	NUMERIC(18, 2)	HODNOTA
ICO	VARCHAR(15)	ICO
FIRMA	VARCHAR(220)	FIRMA
ROK	VARCHAR(4)	ROK
MESIC	VARCHAR(2)	MESIC
DEN	VARCHAR(2)	DEN
DOKLAD	VARCHAR(27)	DOKLAD
RADA	VARCHAR(10)	RADA

Označeno 0 z 8

OK Storno ↻ 🔍

903x375 26 KB

POZNÁMKA

Sloupce ve view nemusejí nutně obsahovat všechny sloupce ze zdrojové tabulky a zároveň nemusejí mít všechny zdrojové tabulky stejnou strukturu, ovšem všechny zdrojové tabulky musejí obsahovat všechny sloupce definované ve view a musejí mít stejný datový typ jako je definován ve view. Také je dobré definovat indexy pro všechny sloupce ve view.