

Kompresia dát trezorov

Kompresia dát trezora je implementovaná na platforme PostgreSQL. Vykonáva optimalizáciu jeho štruktúry a niekoľkokrát zmenší jeho veľkosť. Do komprimovaného trezora už nie je možné zapisovať (ani utilitou `arcsynchro`). Kompresia dát je dostupná v D2000 verzii 21 a vyššej. Nastavenie parametra **TrezorCompress** = 1 zapína kompresiu dát trezora po jeho odpojení.

Aj staršie trezory je možné skomprimovať „rune“ pomocou tell príkazu **TREZOR COMPRESS**. Je podporený aj opaný proces – dekompresia pomocou tell príkazu **TREZOR DECOMPRESS**.

Rýchlosť itania z komprimovaných trezorov je porovnateľná s nekomprimovanými. V niektorých hraniných prípadoch (ítanie konkrétnej hodnoty v aze) môžu byť komprimované trezory pomalšie. V iných prípadoch môže byť ítie komprimovaných dát rýchlejšie (záleží od pomeru rýchlosti a záťaž I/O subsystému a procesora). Navyše kompresia dát umožní uchovávanie vášho množstva dát v pamäti (v cache operaného systému a PostgreSQL).

Prácu s komprimovanými trezormi je možné si vyskúšať „na neisto“. Nastavenie parametra **TrezorCompressKeep** = 1 spôsobí, že pri kompresii vzniknú tabučky s komprimovanými dátami (CDATA, DATA0), ale pôvodná tabučka (DATA) sa nevymaže. Po kompresii sa následne ítajú komprimované dáta. Príkazom **TREZOR DECOMPRESS** je možné sa veľmi rýchlo vrátiť k pôvodnej štruktúre trezora, keďže nie je nutné vykonať dekompresiu, ale staí vymazať tabučky s komprimovanými dátami.

Praktické testy ukázali, že kompresia dát v trezore prebieha rýchlejšie, ak sú predtým dáta upratané. Preto existuje parameter **TrezorCompressReorg**, ktorý je prednastavený na hodnotu 1 a ktorý aktivuje reorganizáciu dáta pred kompresiou.

Výpis trezorov príkazom **LIST_TREZOR** ukazuje pri komprimovaných trezoroch príznak *cps*:

```
Received TELL command : LIST_TREZOR
List all
List of existing trezors:
-----
ID | Data start(UTC) | Data end(UTC) | Status
-----
0134 31-05-2020 22:00 30-06-2020 22:00 being used for writing
0133 30-04-2020 22:00 31-05-2020 22:00 mounted read-only cps
0132 31-03-2020 22:00 30-04-2020 22:00 mounted read-only cps
0131 29-02-2020 23:00 31-03-2020 22:00 mounted read-only
0130 31-01-2020 23:00 29-02-2020 23:00 mounted read-only
```

Obmedzenie PostgreSQL

Komprimované trezory na PostgreSQL podliehajú nasledovnému obmedzeniu: veľkosť dát jedného archívneho objektu (resp. jednej položky štruktúrovaného archívu), ktoré sa ukladajú do jedného riadku komprimovanej dátovej tabučky CDATA, musí byť menšia ako 1 GB. V opanom prípade kompresia skoní s chybou a je vypísané ID objektu, ktorý ju spôsobil. Príklad:

```
[2021-04-05 20:36:52.371]E MES_TREZOR_20 ExecSqlCommand insert into cdata with avgc as (select '20090701
000000.000'::timestamp as cmin, '20090801 000000.000'::timestamp as cmax, '20090701 000000.000'::timestamp +
('20090801 000000.000'::timestamp - '20090701 000000.000'::timestamp)/2 as c), sdata as (SELECT * from data where
"ID"= 1309454 and "ROW"= 0 and "COL"= 0 order by "CAS") select "ID", "ROW", "COL", avgc.c, array_agg( ((extract
(epoch from "CAS"-avgc.c)*1000)::integer, "VALUE", "STATUS", "LIMIT_STATUS", "ARCHIV_STATUS", "FLAGS"))::
d2trzitem) from sdata, avgc where "CAS" between avgc.cmin and avgc.cmax + '1 hour'::interval group by "ID",
"ROW", "COL", avgc.c
(54000)ERROR: array size exceeds the maximum allowed (1073741823);

[2021-04-05 20:36:52.381]E Error inserting cdata for 20 ID=$1309454
```

Popis konfiguraných parametrov kompresie dát trezorov pre platformu PostgreSQL:

- **TrezorCompress** - zapnutie kompresie dát trezora. Kompresia dáta prebehne pri odpájaní trezora. Parameter môže byť zmenený za behu tell príkazom **SET_OPTION TrezorCompress**.
- **TrezorCompressKeep** - parameter slúži na otestovanie kompresie dát trezora. Ak má hodnotu 1, tabučka s pôvodnými dátami (DATA) nebude vymazaná, ale sa ponechá. Dáta sa budú ítať z tabuliek s komprimovanými dátami (CDATA, DATA0). Príkazom **TREZOR DECOMPRESS** je možné sa veľmi rýchlo vrátiť k pôvodnej štruktúre trezora, keďže nie je nutné vykonať dekompresiu, ale staí vymazať tabučky s komprimovanými dátami. Parameter môže byť zmenený za behu tell príkazom **SET_OPTION TrezorCompressKeep**.

Ak chcete **vymazať komprimované dáta a zachovať nekomprimované**, použite nasledovný postup:

- Príkazom **SET_OPTION TREZOR_COMPRESS_KEEP OFF** nastavte hodnotu parametra **TrezorCompressKeep** na 0.

- Príkazom **TREZOR DECOMPRESS** vykonajte dekompresiu dát trezora (dekompresia prebehne rýchlo, keďže budú iba vymazané tabučky s komprimovanými dátami - CDATA, DATA0).

Ak chcete **zachovať komprimované dáta a vymazať nekomprimované**, použite nasledovný postup:

- Príkazom **TREZOR DECOMPRESS** vykonajte dekompresiu dát trezora (dekompresia prebehne rýchlo, keďže bude iba trezor označený ako dekomprimovaný. Vďaka hodnote parametra **TrezorCompressKeep=1** nebudú vymazané tabučky s komprimovanými dátami - CDATE, DATA0).

- Príkazom **SET_OPTION TREZOR_COMPRESS_KEEP OFF** nastavte hodnotu parametra **TrezorCompressKeep** na 0.

- Príkazom **TREZOR COMPRESS** vykonajte kompresiu dát trezora (kompresia prebehne rýchlo, keďže sa použijú existujúce dáta v tabučkách s komprimovanými dátami). Kompresia vymaže tabučku s pôvodnými nekomprimovanými dátami (DATA).

- **TrezorCompressReorg** - reorganizácia dát trezora pred kompresiou. Tento parameter má prednastavenú hodnotu 1 - v praxi sa ukazuje, že je efektívnejšie a rýchlejšie najskôr zreorganizovať dáta (SQL príkazom *CLUSTER DATA*) a až následne vykonať ich kompresiu. Hodnota 0 vypína reorganizáciu dát trezora pred kompresiou. Vypnutie reorganizácie pred kompresiou nie je odporúčané. Parameter môže byť zmenený za behu tel príkazom **SET_OPTION TrezorCompressReorg**.
- **TrezorCompressCmt** - parameter udáva počet archívnych objektov, po ktorých komprimovaní sa vykoná COMMIT. Prednastavená hodnota je 10. V aplikáciách s intenzívne archivovanými objektami (veľa hodnôt jedného archívneho objektu v trezore) je možné tento parameter znížiť až na hodnotu 1. Naopak, v aplikáciách s veľa archívnych objektami, ktoré sa menia zriedkavo, je možné hodnotu tohto parametra zvýšiť. Parameter môže byť zmenený za behu tel príkazom **SET_OPTION TrezorCompressCmt**.
- **TrezorCompressOrder** - parameter sa týka itania z trezorov s komprimovanými dátami. Tento parameter udáva, i pri itaní dát z komprimovaného trezora je nutné použiť zoradenie podľa asu. Keďže sú dáta zoradené pri kompresii, štandardne je možné nechať parameter na hodnotu 0, o zrychľuje itanie z komprimovaného trezora (vynecháva sa ORDER BY). Parameter môže byť zmenený za behu tel príkazom **SET_OPTION TrezorCompressOrder**.

Postup kompresie starých trezorov

- Ak boli staré trezory už upratané (napr. zálohovací skript obsahoval "cluster data"), je možné pred kompresiou vypnúť upratovanie tel príkazom: **SET_OPTION TREZOR_COMPRESS_REORG OFF**
- Trezory, ktoré budú komprimované, treba pripojiť na zápis. Najskôr ich teda treba odpojiť a následne pripojiť. Napríklad pre trezory 1-10:
DISMOUNT_TREZOR 1 10
MOUNT_TREZOR 1 10 WRITE
- Teraz je možné trezory komprimovať tel príkazom **TREZOR COMPRESS <trezor_id>**
napr. **TREZOR_COMPRESS 1**
Trezor s veľkosťou okolo 20 GB sa na štandardnom serveri komprimuje okolo hodiny - dvoch. Ak je to oveľa viac, tak zrejme nie sú dáta "upratané" a treba zapnúť upratovanie pred kompresiou (**SET_OPTION TREZOR_COMPRESS_REORG ON**).
Štandardne má archív zapnutú debug kategóriu *DBG.ARCHIV.COMPRESS.TREZOR* a v D2000 Sysconsole sú viditeľné debug výpisy. Hovorí o kompresii konkr. objektov (resp. položiek štruktúrovaných archívov), v zátvorke je poradové číslo / celkový počet objektov). Raz za 10 objektov nasleduje commit do databázy (toto sa dá parametrizovať tel príkazom **SET_OPTION TREZOR_COMPRESS_CMT <počet>**).
Compress trezor 81 ID \$11291(1048/ 39113)
Compress trezor 81 ID \$11291(1048/ 39113) done
Compress trezor 81 ID \$11293(1049/ 39113)
Compress trezor 81 ID \$11293(1049/ 39113) done
Compress trezor 81 ID \$11295(1050/ 39113)
Compress trezor 81 ID \$11295(1050/ 39113) done (commit)
Compress trezor 81 ID \$11297(1051/ 39113)
Compress trezor 81 ID \$11297(1051/ 39113) done
Compress trezor 81 ID \$11299(1052/ 39113)
Compress trezor 81 ID \$11299(1052/ 39113) done
- Po skomprimovaní je vhodné trezory odpojiť a pripojiť na itanie. Napríklad pre trezory 1-10:
DISMOUNT_TREZOR 1 10
MOUNT_TREZOR 1 10
- Tel príkazom **LIST_TREZOR** je možné skontrolovať, že sú skomprimované všetky požadované trezory. Skomprimovaný trezor má príznačenie "cps".
Napríklad trezor 140:
0140 31-07-2019 22:00 31-08-2019 22:00 mounted read-only cps
- Ak bola vypnutá kompresia pred upratovaním, odporúčame ju zapnúť tel príkazom: **SET_OPTION TREZOR_COMPRESS_REORG OFF**
- Kompresia novovytvorených trezorov sa zapne tel príkazom: **SET_OPTION TREZOR_COMPRESS ON**