

Kada brzo nije dovoljno

Jedan od načina kako se može dobiti veliko ubrzanje obrade informacija jest prebacivanje što veće količine podataka u bržu memoriju

Mario Jurić, Technology & Development Manager, Megatrend poslovna rješenja

Svi već znaju koliko su informacije bitne i kako trenutačno živimo u svijetu u kojem se količina podataka svakodnevno povećava. U stalnoj utrci s vremenom, vrlo je bitno da su podaci dostupni na vrijeme i da informacije koje tražimo dobijemo što prije – prema mogućnostima, trenutačno.

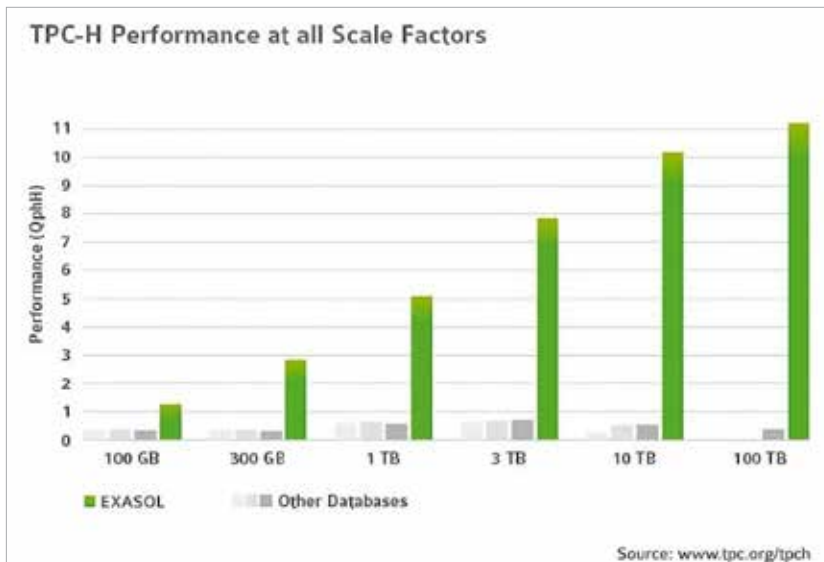
Zbog tog zahtijeva tradicionalne relacijske baze podataka često nisu dovoljno brze, i kako bismo dobili zadovoljavajuće performanse moramo koristiti namjenske baze za analitiku. Jedan od načina kako se može dobiti veliko ubrzanje jest prebacivanje što veće količine podataka u bržu memoriju. Smanjenje cijena radne memorije i povećanje njezina kapaciteta dovelo je do razvoja *in-memory* baza podataka i njihove sve veće zastupljenosti.

NAJBRŽA IN-MEMORY BAZA

Primjer takve baze je Exasol, trenutačno najbrža *in-memory* analitička baza na svijetu. Kako bi se povećala iskoristivost hardvera, koristi se paralelna obrada i komunikacijske tehnike visokih performansi na različitim razinama arhitekture. Te su tehnike namjenski dizajnirane i izgrađene od nule za Exasol bazu, kako bi postigle vrhunske performanse bez ograničenja volumena podataka.

To je potvrđeno i na TPC-H testiranju, koje se smatra industrijskim standardom na području analitičkih baza. Exasol već godinama drži uvjerljivo prvo mjesto ispred konkurencije u svim kategorijama (od 100 GB do 100 TB obradenih podataka), i to ne samo u performansama, već i u omjeru performanse/cijena rješenja. Dodatna korist je to što osim brzine, omogućuje korisnicima pokretanje većih i kompleksnijih analitičkih upita i korištenje baze u mnogo širem rasponu primjena. Često postoji pogrešna predodžba o *in-memory* analitičkim bazama podataka da one za učinkovit rad sve trajne i privremene podatke moraju držati u radnoj memoriji sve vrijeme, što bi osjetno ograničavalo količinu podataka koja se može pohraniti i obrađivati u takvim sustavima. Za Exasol, takva ograničenja nisu problem.

Nakon više od dva desetljeća istraživanja i ulaganja u *in-memory* tehnologiju, Exasol je dosegno visoku razinu zrelosti i svestranost te je dobio dublje razumijevanje onoga što *in-memory* tehnologija zapravo jest. Ona



Rezultat
TPC-H
testa



Exasol kao srce analitičkog okruženja

brzo

nije samo jedna tehnička značajka sustava, već cjelokupni dizajn pristupa za pohranu podataka i obradu algoritama, pristup koji radi na cijelom sustavu. Zbog toga je Exasol vrlo dobar u predviđanju budućih pristupa podacima, što omogućuje bolje upravljanje podacima na svim razinama granularnosti, bilo da se radi o priručnoj memoriji, blokovima u radnoj memoriji, ili podacima koji se repliciraju s udaljenih strojeva. Bitnu ulogu u učinkovitosti imaju pohrana i sažimanje podataka u stupcima, što omogućuje i do tri puta više sirovih podataka koji stanu u radnu memoriju te manje opterećenje na priručne memorije. Također, to omogućuje i smanjenje broja ulazne i izlazne operacije prilikom čitanja i zapisivanja podataka.

MASIVNO PARALELNO PROCESIRANJE

Dodatno ubrzanje dobiva se masivnom paralelnom obradom (engl. Massively Parallel Processing - MPP) koja omogućuje usporedni rad više procesora ili čvorova na velikoj količini podataka. Na razini klastera Exasol prati SPMD (single program, multiple data) paradigmu, što znači da različiti čvorovi u klasteru izvršavaju isti program (upit) nad različitim podacima, a za vrijeme izvođenja njihovo je interno stanje veći dio vremena neovisno od drugim čvorovima.

Podrška za masivnu paralelnu obradu omogućuje i jednostavno horizontalno skaliranje dodavanjem dodatnih čvorova u kla-


ster. I s financijske strane bitno je je li sustav horizontalno skalabilan i je li moguće jednostavno povećati njegov kapacitet i performanse dodavanjem dodatnih servera. Ako to nije slučaj, to znači i dodatne, poprilično visoke, troškove u slučaju potrebe za nadogradnjom koja će se kad-tad pojaviti.

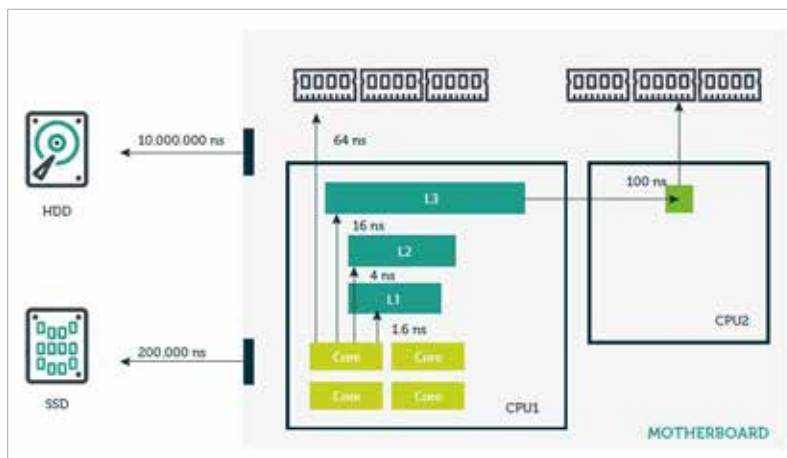
JEDNOSTAVNOST KORIŠTENJA

Exasol je samooptimirajuća baza, što znači da se puno optimizacijskog posla odrađuje bez ljudske interakcije. Primjerice, indeksi koji se koriste za mnoge operacije, poput spajanja tablica ili filtriranja, kreiraju se i održavaju automatski, bez zamaranja korisnika takvim detaljima. U praksi to znači veliku uštedu vremena koje se inače utroši na održavanje i optimizaciju baze.

S obzirom na to da je usklađena sa standardom ANSI SQL te kako podržava različita sučelja poput JDBC, ODBC i ADO.NET, jednostavno se integrira u postojeću analitičku arhitekturu. Za početak to može biti samo

preuzimanje kritičnih poslova koji su se na postojećem sustavu sporo izvršavali, poput dugotrajnih složenih obrada, ili kao brzi memorijski sloj za ključne poslovne analitike. Vrijednost koju donosi Exasol prepoznata je i u izvještaju Dresner Advisory Services: Analytical Data Infrastructure Market Study 2018, u sklopu kojeg je Exasol identificiran kao lider u kategorijama vjerodostojnosti dobavljača i iskustvu korisnika.

Dakle, nemojte čekati – već jednostavno isprobajte. Baza je dostupna u obliku softvera koji možete instalirati na standardnu x86 arhitekturu, kao namjenski uređaj koji uključuje i hardver i predinstalirane softwere, ili kao servis u *cloudu* (Amazon, EXACloud, Azure). Postoji i besplatna verzija Exasol Community Edition, koja nudi sve funkcionalnosti, uz ograničenje od maksimalno 1 TB podataka, i instalaciju na samo jedan čvor. Za dodatne informacije i stručne savjete slobodno se obratite našim stručnjacima ili na www.megatrend.com. 



Vrijeme pristupa različitim vrstama memorija

EXASOL: 10 KLJUČNIH ZNAČAJKI

1. In-memory analitika	Inovativni <i>in-memory</i> algoritmi omogućuju obradu ogromnih količina podataka u radnoj memoriji, uz značajno brže vrijeme pristupa
2. Masivna paralelna obrada	Baza je razvijena kako paralelni sustav, temeljen na "shared nothing" arhitekturi
3. Sažimanje i spremanje podataka u stupcima	Smanjuje broj ulazno/izlaznih operacija i količinu podataka potrebnu za obradu u radnoj memoriji i ubrzava performanse
4. Skalabilnost	Linearna skalabilnost omogućuje nadogradnju sustava i povećanje performansi dodavanjem dodatnih čvorova
5. Samopodešavajuća baza	Inteligentni algoritmi prate korištenje baze i odrađuju samopodešavanje, što rezultira optimalnim performansama i minimizira suvišnu administraciju
6. Brži pristup različitim izvorima podataka	Virtualne sheme, kao i integracijsko sučelje visokih performansi, omogućuju povezivanje i analizu podataka iz različitih izvora
7. Podrška za Hadoop	Podrška za sve izvorne HDFS formate omogućuje izvršavanje brzih analitičkih upita na strukturiranim i nestrukturiranim podacima
8. Napredna analitika unutar baze podataka	Podrška za programske jezike R, Python, Java i Lua kao i integraciju bilo kojeg drugog programskog jezika te njegovo korištenje za analitiku unutar baze podataka
9. Nenadmašna povezivost	Jednostavno povezivanje s postojećim BI i integracijskim alatima preko ODBC, JDBC, ADO.NET sučelja, kao i na JSON-u baziranih web-socket API-ja
10. Ultimativna fleksibilnost	Exasol se uklapa u vaš poslovni model i može se koristiti na različite načine samo kao softver, namjenski uređaj (appliance) ili u <i>cloudu</i>