

# Samooptimizirajuća ana baza podataka

**Zahtjevi današnjeg tržišta nalažu svakoj tvrtki da se sve više usredotoči na vlastite podatke i podatke klijenata – zapravo, na sve podatke, neovisno o njihovom izvoru, bilo u obliku zasebnih datoteka, organiziranih sustava senzora u proizvodnji ili postojećih skladišta podataka i internog softvera koji koristi za svakodnevno poslovanje**

Vedran Podubski, konzultant za poslovna rješenja, Megatrend poslovna rješenja

**K**oličina podataka za samu obradu, bilo analitičku ili statističku, za predviđanje budućeg poslovanja i njegovu optimizaciju te svakodnevni rad unutar tvrtke – sve je veća, a time i potreba za što većom brzinom obrade, pa se i samo “skupljanje” podataka na jednom mjestu eksponencijalno produljuje. Često viđen slučaj u praksi je potreba za ulaganjem velike količine materijalnih resursa u modernizaciju postojećih sustava, u obliku kupnje novog, dodatnog, hardvera (novi serveri, novi procesori, što veća količina radne memorije) te zapošljavanja dodatnog, često i skupog kadra, kako bi se riješili problemi s usporavanjem rada digitalnih skladišta podataka.

Takve investicije često se pokazuju i neuspješnima zbog povećanja složenosti nadogradnje postojeće infrastrukture (nekompatibilnosti s postojećim sustavima na razini hardvera, duljih razdoblja zastoja u radu produkcijskih baza, itd.) i dodatno povećavaju već visoke troškove poslovanja. Kao popularnom rješenju takvih problema, sve više tvrtki u svijetu pa i u regiji, okreće se samooptimizirajućim analitičkim *in-memory* bazama podataka koje ne zahtijevaju “duboko” interno podešavanje, nego imaju razvijene napredne interne algoritme koji optimiziraju svaki dio baze te automatski i bez ljudske intervencije drastično ubrzavaju obradu podataka.

Uz to, imaju i lako skalabilnu arhitekturu koja se može proširiti ili smanjiti, bez potrebe za prekidanjem rada produkcijskih baza



Struktura Exasol baze podataka

podataka. Navedene baze podataka mogu se koristiti kao primarno skladište podataka ili kao “međusloj” koji služi za ubrzanje postojeće digitalne infrastrukture analitičkih i izvještajnih sustava, a koštaju osjetno manje od “tradicionalnih” RDBMS sustava. Trenutačno najbrža i najpovoljnija takva baza podataka na tržištu je Exasol.

## INTELENTNI ALGORITMI

Exasol je *in-memory* baza podataka, objavljena 2000. godine, nakon uspješne implementacije naprednih algoritama razvijenih u razdoblju od 1990. do 2000. godine na Sveučilištu Friedrich-Alexander na fakultetima u Erlangenu i Nürnbergu u

Njemačkoj. Od 2008. godine, Exasol je jedan od potencijalnih favorita među platformama koje kao baze podataka služe za analitiku i izvještavanje. Na samom je vrhu TPC-H testiranja u području analitičkih baza podataka, a navedeni rezultati i dalje se održavaju u kategorijama od 1000 GB + obradenih podataka. Exasol, trenutačno, ima i najbolji omjer performanse i cijene rješenja na tržištu, prema podacima s kraja srpnja 2019. godine. Stalnim razvojem i unaprjeđivanjem postojećih algoritama i načina zapisivanja, čitanja i obrade podataka unutar same baze podataka već više od dva desetljeća, dobiva se iznimna brzina i iskoristivost hardvera na

različitim razinama arhitekture u svakodnevnom poslovanju s bazom. Exasol je samooptimizirajuća baza podataka koja većinu prilagodbi odraduje sama, pomoću internih inteligentnih algoritama, bez potrebe za ljudskom intervencijom i interakcijom.

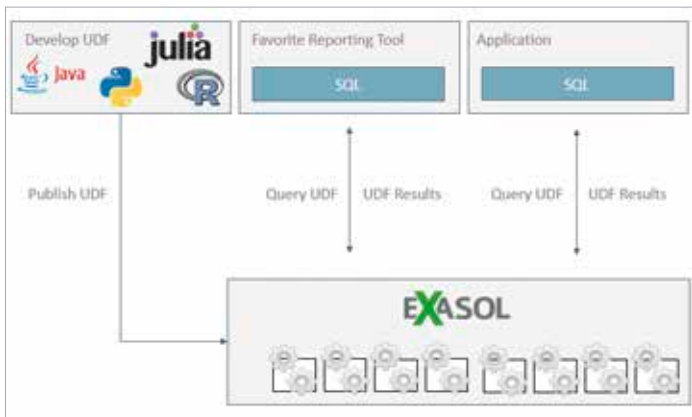
Optimizacija indeksa u svakodnevnom operacijama kao što su filtriranje, spajanje tablica, izvođenje kompleksnih upita, potpuno je automatizirana te ujedno razvijena tako da predviđa sljedeće moguće upite na temelju povijesnog korištenja baze, kako bi se njena optimizacija odradila i prije izvršenog zahtjeva, posebno kod složenih analitičkih i/ili izvještajnih poslova. To čini samu Exasol tehnologiju bržom od konkurencije. Uz automatsku optimizaciju indeksa, interni sustav u bazi predviđa i budući pristup podacima, također na temelju povijesnog korištenja na svim razinama; u priručnoj memoriji, blokovima, radnoj memoriji ili u dodatnim serverima/čvorovima u clusteru.

## UBRZANA OBRADA PODATAKA

Učinkovitost i brzina obrade podataka kod Exasola se ostvaruje primjenom nekoliko osnovnih načela funkcioniranja; prvo na načelo je usporedna obrada golemih količina podataka (engl. Massively Parallel Processing – MPP) na više procesora ili čvorova. Kod cluster sustava, različiti čvorovi izvršavaju isti upit nad različitim skupovima podataka, neovisno o međusobnom stanju čvorova. Preciznije rečeno, primjenjuju paradigmu SPMD (single program, multiple data), komunicirajući međusobno preko interne rezervirane mreže, paralelno, principom “shared-nothing” arhitekture.

Zbog samog funkcioniranja obrade podataka na MPP principu olakšava se i horizontalno skaliranje, dodavanjem dodatnih čvorova u cluster, što sustav čini skalabilnim, bez potrebe za kompleksnim migracijama podataka i zastojem u radu same baze, jer se dodatni čvor jednostavno instalira i skalira kroz samo administrativno sučelje, automatskim procesom ugrađenim

# litička in-memory



Primjer podržanih korisnički definiranih funkcija

## SPECIJALIZIRANE SKRIPTE

Sama Exasol baza potpuno je "ACID compliant" (engl. Atomicity, Consistency, Isolation, Durability) te od verzije 6 uključuje i mogućnost izvršavanja UDF (engl. User Defined Function) skripti za dodatnu proširenu analitiku podataka unutar baze (engl. Extendable Data Analytics Framework), podržanih u više programskih jezika kao što su, primjerice, R, Python, Lua, C++, s mogućnošću njihova pozivanja iz vanjskih aplikacija za izvještaje, primjerice, Tableau, Qlik, Rstudio. Sama baza u potpunosti je usklađena s ANSI SQL standardom te podržava različita sučelja za ostvarivanje konekcije poput ODBC, JDBC, .NET, a web API je na bazi jezika JSON.

Jednostavno se integrira u postojeću analitičku arhitekturu. Za početak, to može biti tek preuzimanje

kritičnih poslova koji su se na postojećem sustavu sporo izvršavali - poput dugotrajnih složenih obrada, ili može poslužiti kao brzi memorijski sloj za ključne poslovne analitike. Exasol nudi i

razne mogućnosti održavanja kontinuiteta rada kod mogućih problema s hardverom (Disaster Recovery), uključujući i mogućnosti replikacije u više podatkovnih centara te sinkronu i asinkronu replikaciju u *online* i *standby* režimima. ◀

Rank	Company	System	QphH	Price / QphH	Watts / KQphH	System Availability	Database	Operating System	Date Submitted	Cluster
1		Dell PowerEdge R6415	6.053.020	.07 USD	NK	07/09/19	EXASOL 6.2	CentOS 7.6	07/09/19	Y

Rezultat TPC-H testa do 3000GB



u sustav. Sljedeći vrlo bitan faktor kojim se ostvaruje ubrzanje rada je tzv. pohrana i sažimanje podataka u stupcima (engl. Column-based storage and compression) kojom

se smanjuje potrebni broj ulaznih i izlaznih (I/O) operacija (čitanje i pisanje na diskovni/memorijski sustav) i do tri puta, čime se omogućuje obrada veće količine poda-

taka u radnoj memoriji. Pohrana i sažimanje podataka u stupcima, razvijena od Exasola, povećava učinkovitost same baze, jer se radni i privremeni podaci mogu brže zapisivati, i u većoj količini, iz radne memorije na disk i obrnuto, uzevši u obzir sve do sada navedene parametre funkcioniranja.

## APLIKACIJA ILI UREĐAJ ILI SERVIS ILI CLOUD

Exasol bazu u posljednjih su nekoliko godina uspješno u svoje sustave integrirale velike svjetske kompanije kao što su Adidas, Badoo, XING te tvrtke u regiji kao što je Telekom Slovenije. Prema testiranjima u regiji - izvršenima sredinom 2019. godine - na podacima veličine oko 400 GB, procedure za generiranje izvještaja nad Exasol bazom ostvarile su ubrzanje od 10 do 30 puta naspram konkurentnih baza podataka.

Uz već navedeni rezultat testiranja od TPC-a, nezavisna testiranja brzine, primjerice od ruske banke Tinkoff Bank, provedenih u siječnju ove godine, u usporedbi s glavnim konkurentskim bazama podataka, Exasol pokazuje osjetno manja vremena od svih konkurenata. Baza je dostupna u obliku softvera koji možete instalirati na standardnu x86 arhitekturu, kao namjenski uređaj koji uključuje i hardver i predinstalirane softvere, kao servis u *cloudu* (Amazon, EXAcloud, Azure, Google Cloud) ili nešto kompleksniji, hibridni *cloud*. Postoji i besplatna verzija Exasol Community Edition, koja nudi sve funkcionalnosti i instalaciju na samo jedan čvor. Za dodatne informacije i stručne savjete oko Exasol baze slobodno se obratite našim stručnjacima na [poslovna\\_rjesenja@megatrend.com](mailto:poslovna_rjesenja@megatrend.com).