

# Baza koja brže obrađuje goleme količine podataka

**Digitalizacija poslovanja**, potaknuta rastućim tehnološkim mogućnostima, u fokusu je velikog broja tvrtki. Količina podataka i brzina njihove obrade od velikog je utjecaja na uspjeh spomenutih projekata i procesa



Vedran Podubski, konzultant za poslovna rješenja, Megatrend poslovna rješenja

“Tradicionalne” baze podataka, osim što generiraju velike troškove (zbog spore obrade podataka troše se dodatni sati za optimizaciju, ulaže se u jači-brži-bolji hardver, a sve u svrhu bolje radne učinkovitosti, nisu prikladne za današnje potrebe, odnosno ne mogu zadovoljiti sve veće i veće zahtjeve poslovnih korisnika.

## UBRZAVANJE ANALITIČKIH OPERACIJA

Izvođenje analitičkih operacija nad skupom digitalnih podataka

unutar tvrtke - neovisno o tome je li riječ o analizi povijesnih podataka, nadzoru poslovanja ili, pak, predviđanjima, odnosno planiranju budućnosti - ne bi smjelo biti točka zagušenja, pa mnoge tvrtke istražuju moguća rješenja, odnosno integriraju u svoje sustave suvremene baze podataka, napravljene specijalno sa svrhom ubrzanja analitičkih operacija nad velikim količinama podataka. U tu svrhu, istražujući i testirajući aktualne baze, odlučili smo koristiti i preporučivati svojim klijen-

tima modernu, pouzdanu i vrlo brzu bazu podataka - Scream DB. Scream DB je RDBMS baza, objavljena još 2014. godine, osmišljena za skladištenje velikih količina podataka (Big Data) koja se koristi dodatnim ubrzanjem primjenom grafičkih procesora (GPU). Scream DB baza napravljena je od početka, odnosno niti jedan prethodno postojeći sustav nije korišten kao temelj za njen razvoj. Izvršavanje upita na grafičkim procesorima tehnologija je slična sustavima korištenim

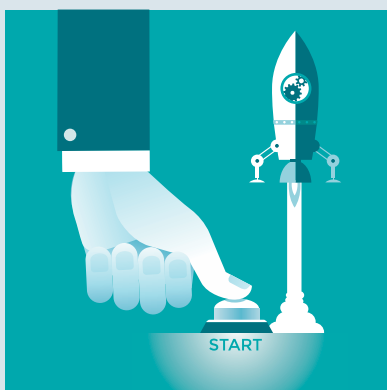
za rudarenje podataka kod kriptovaluta te omogućuje masovno paralelno procesiranje podataka na svakoj jezgri procesora grafičke kartice, koristeći bržu frekvenciju grafičke memorije od standardne RAM memorije na matičnoj ploči.

## HORIZONTALNO I VERTIKALNO PARTICIONIRANJE

U standardnim skladištima podataka sve komponente unutar sustava usko su povezane i zajednički koriste sve hardverske resurse te je pri velikom protoku podataka i velikom broju korisnika, skaliranje otežano i stvaraju se problemi s performansama. Scream DB rješava taj problem inteligentnom internom arhitekturom, koristeći odvojeni kompajler, izvršni dio i spremnik podataka te time optimizira protok podataka i njihovu obradu. Dodatno poboljšanje performansi Scream DB postiže se potpuno automatskim particioniranjem podataka na dva načina (hiperparticioniranje). Prvi dio particioniranja je vertikalni (kolumarni), i omogućuje selektivni pristup podskupovima kolona u bazi, čime smanjuje potrebu za čestim pisanjem/čitanjem s diskova, čineći ga savršenim za paralelnu obradu podataka korištenjem grafičkog procesora.

Drugi dio particioniranja je horizontalni, uz podjelu na komade i opsege (chunks and extents). Horizontalna podjela podataka na

## SQREAM ZA GURUE



Za dodatne informacije i stručne savjete oko Scream DB i/ili IBM POWER9 sustava slobodno se obratite našim stručnjacima na poslovna.rjesenja@megatrend.com

### 1. SQREAM DB METODA

Scream DB je potpuno ANSI - 92 SQL kompatibilna i lako se implementira u sve ekosustave zbog podrške za sve tipične ODBC i JDBC konektore, uključujući i Python, C#, .NET, C++, Java i druge. Nativna podrška za SQL jezik omogućuje korištenje bilo kojeg ETL alata i ostalih aplikacija nad bazom, smanjujući vrijeme implementacije na minimum.



## KAD SPARK I HADOOP NISU DOVOLJNO BRZI...

...ili kako smo riješili izazov (da ga ne bismo zvali problemom :), kod klijenta iz finansijskog sektora

### IZAZOV

U finansijskom sektoru povjerenje klijenata, kako građana tako i poslovnih korisnika, odlučujuće je. S ciljem zaštite svojih klijenata i njihovih podataka od prijevара, DS tim (Data Science team) treba provoditi dugotrajna istraživanja i analitiku na ogromnoj količini povijesnih podataka. DS tim ima ozbiljnih problema u pripremi podataka, u procesu generiranja novih svojstava AI modela, odnosno općenito u raznim dijelovima AI procesa, koje obrađuju na dijeljenoj Hadoop infrastrukturi. Upiti koji se dugo izvršavaju često se prekidaju ili ih deprioritiziraju drugi korisnici Hadoop sustava. Uz to, neki se komplicirani upiti uopće ne mogu provesti na postojećoj infrastrukturi. Tražilo se rješenje koje će uspješno riješiti navedene probleme i izazove, uz zadržavanje fokusa na održavanju povjerenja klijenata prilikom korištenja naprednih tehnika otkrivanja prijevара (fraud detection).

### RJEŠENJE

Nakon opsežnog testiranja nekoliko rješenja različitih proizvođača, DS tim je odlučio integrirati SQream bazu u postojeću infrastrukturu baziranu na Sparku i Hadoopu kako bi osnažio ubranu analitiku velikih količina podataka. Nova arhitektura *offloada* važne podatke na SQream DB server, omogućujući DS timu da provodi napredne ad hoc upite, priprema podatke i generira veliki broj svojstava kako bi nahranio AI *pipeline*, i sve to u kratkom vremenu i uz mali dio korištenih infrastrukturnih resursa.

### UČINAK

Kako se novo rješenje odrazilo na poslove DS tima? Komplicirane funkcije upita koje pretražuju stotine tisuća prodavača, desetke milijuna klijenata i milijarde transakcija, izvršavaju se u manje od sat vremena (prethodni sustav za to je trebao više sati). DS tim sada može izvršavati kompleksne upite koje prije, koristeći samo Hadoop cluster, nije mogao. Postojanje dedicanog sustava za DS tim povećalo je i učinkovitost i neovisnost DS tima o drugim odjelima. SQream DB koristi mnogo manju HW infrastrukturu (u konkretnom

slučaju x86 servera s 4 NVIDIA Tesla V100 16GB GPU povezana sa sustavom za pohranu podataka), i to uz mnogo bolje performanse za napredne upite. Učitavanje (load) 20 je puta brže, uz 30% manje korištenja diskovnog prostora u odnosu na prijašnje korištenje Sparka i Hadoopa. Uz SQream bazu na raspolaganju, DS tim sada se može fokusirati više na analitiku, a manje na borbu s resursima, samim time osiguravajući bolju uslugu internim korisnicima, i to i bržim isporukama traženih analiza, ali i novim uvidima. DS tim je proširio opseg proučavanja podataka te sada obrađuju nekoliko godina povijesnih podataka (u odnosu na nekoliko mjeseci povijesnih podataka koje su obrađivali prije), a što je rezultiralo boljim i preciznijim modelima predviđanja prijevара i zaštite klijenata.

### O KLIJENTU

Međunarodna finansijska institucija provodi više od 10.000 transakcija svake minute. Od samih početaka kad je, tijekom 60-ih, izdala prvu kreditnu karticu, tvrtka slovi kao pouzdana, inovativna te kao tržišni lider desecima milijuna klijenata po cijelom svijetu.

### IZJAVA KLIJENTA

Kontinuirano istražujemo nove i bolje načine analize golemih količina podataka. SQream baza omogućuje našem DS timu da provodi uvide koji su nam potrebni, i sve to uz malu nadogradnju na naš postojeći Hadoop sustav. SQream DB projektirana je da koristi GPU pri izgradnji skladišta podataka. Njen specifičan dizajn omogućuje analitiku i obradu ogromnih količina podataka te je stoga tvrtke koriste za poboljšanje performansi, smanjeni *footprint*, zamjetnu uštedu, odnosno smanjenje troškova, te mogućnost skaliranja količine podataka koje obrađuju sve do stotina terabajta i više.

Megatrend poslovna rješenja ovlaštenu su partner tvrtke SQream Technologies te klijentima, identifikacijom problema, postavljanjem POC-a, testiranjem predloženog rješenja i samom implementacijom i održavanjem, osiguravaju uspješnije poslovanje.

manje podskupove omogućuje bolje iskorištenje hardvera i relativno male količine GRAM-a (RAM na grafičkoj kartici) inteligentnim ko-

rištenjem predmemorije (cache) i spajanja podataka (engl. spooling). Standardne baze podataka koje se koriste u skladištima podataka

koriste samo procesorske jezgre i RAM memoriju za obradu prilikom upisivanja i dohvaćanja podataka. Kod SQream DB taj je

proces proširen na inteligentno korištenje kombinacije dostupnih resursa procesora, RAM memorije i grafičkih procesora. M

## 2. SQREAM DB EKOSUSTAV

SQream DB može raditi na većini standardnog serverskog x86 - 64 hardvera s Nvidijinim grafičkim karticama, pa čak i komercijalnim laptopima opremljenim takvim hardverom, no za najbolje performanse preporučuju se grafičke kartice 2x Nvidia Tesla (K80, P40, P100 itd.) te procesori IBM POWER9.



## 3. IBM POWER9 VS. INTEL

Baza je dostupna u obliku softvera koji možete instalirati na standardnu x86 - 64 ili IBM POWER9 arhitekturu s grafičkim karticama NVIDIA, kao servis u *cloudu* (Amazon P2 / P3 with NVIDIA Tesla, Azure NCv3 with Tesla V100) i IBM Bluemix bare-metal sustave.

