

Akceleratori baza podataka kolumnarne arhitekture

Goran Cvijanović

Vinteh d.o.o.

listopad 2010



Sadržaj prezentacije

- Baze podataka ROW vs COLUMN
- Zašto i kada koristiti baze s kolumnarnom strukturom
- Baze za poslovnu inteligenciju
- InfiniDB, Infobright
- Platforme za BI koje mogu koristiti kolumnarne baze
- Performanse i skalabilnost
- Zaključak

Open Source advokatura



Copyright is
for losers©™

- NO COPYRIGHTS OR LICENSES. DO WHAT YOU LIKE.

This is the new jQuery Tools license

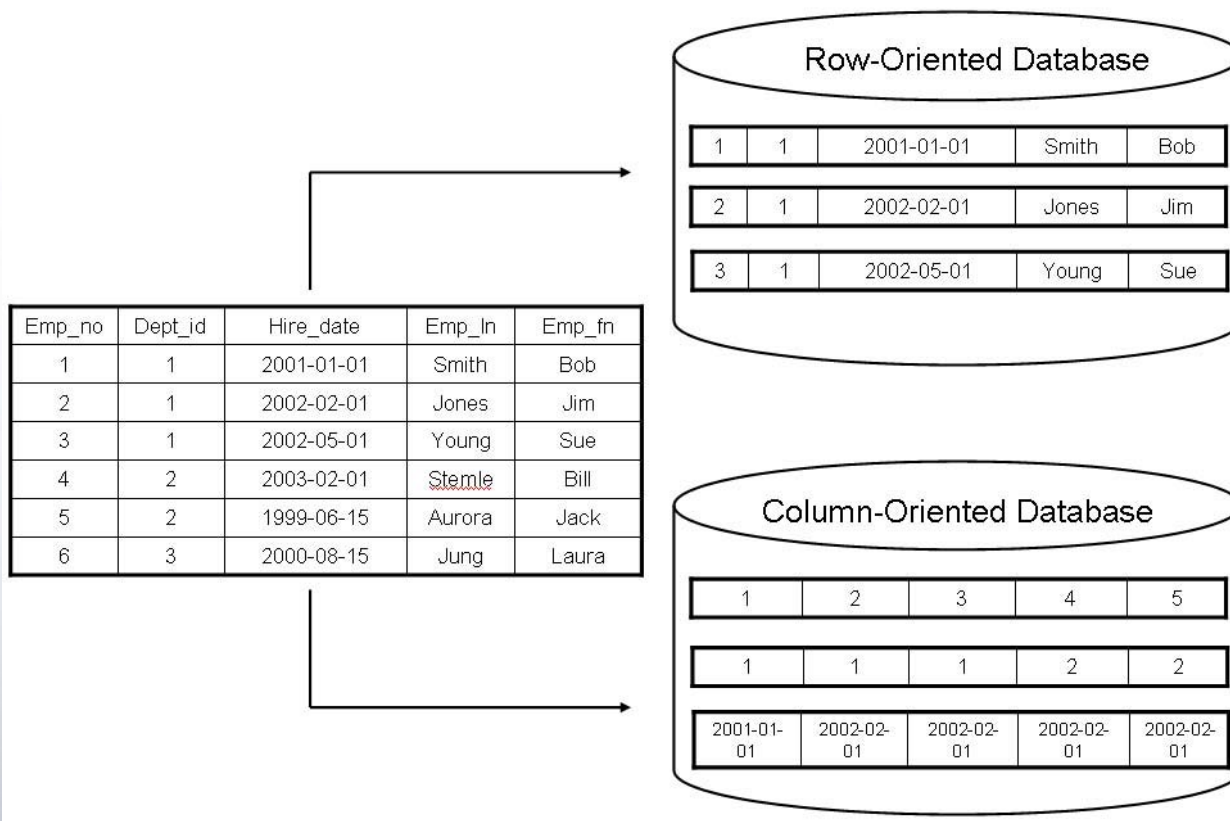
Copyrights and patents are evil. They block the natural progress of development. We all know it - if people start sharing instead of owning the world would be a better place

Open Source advokatura - g.c.



- Optimiranje baza
- BI expertiza
 - integracija podataka
 - Pentaho i Palo platforme
 - izgradnja skladišta podataka
- Izgradnja BI rješenja za velike količine podataka i broj korisnika
- Virtualizacija

Strukture baza podataka



- Kolumnarne baze spremaju podatke na drugačiji način od klasičnih relacijskih baza
 - ROW
 - OLTP baze
 - COLUMN
 - OLAP baze
- [ali ne nužno]

Prednosti kolumnarne arhitekture



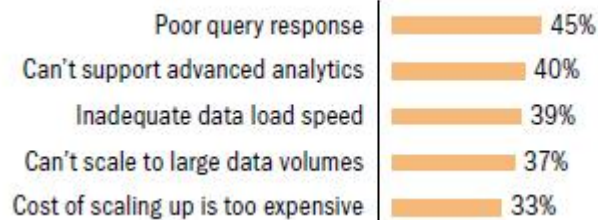
- Kolumnarne baze dohvaćaju samo one kolone koje su navedene u upitu
- Efekti selektivnog dohvata zamjenjuju potrebu za indeksima
- Za tablice sa velikim brojem kolona značajno poboljšava performanse
- Smanjuje broj I/O upita do 75%, te eliminira potrebu za particioniranjem
- Uobičajeno ubrzavaju izvršavanje upita za 5 do 10 puta
- Nepotrebni su materijalizirani pogledi (view) i prekalkulirane vrijednosti

Kada k.a. nisu prednost



- Ukoliko aplikacija koristi puno SELECT * upita, dohvaćaju se sve kolone
- Kada se izvršava veliki broj pojedinačnih INSERT i DELETE naredbi koje su u načinu rada zapravo RAW orjentirane (za razliku od UPDATE)
- Za korištenje u sustavima koji su prvenstveno OLTP
- Za male baze podataka, dobitak na performansama je manji

Trendovi u BI sferi



TDWI, 2009

- Razlozi za korištenje novih arhitektura baza podataka za BI
 - loše performanse upita 45%
 - napredne analitičke funkcije nisu podržane 40%
 - nedovoljna brzina učitavanja podataka 39%
 - nemogućnost skaliranja za velike količine podataka 37%
 - cijena licenci i ostalih troškova mogućeg skaliranja je previsoka 33%

Predviđanja analitičkih kuća



- Most data warehouses will be stored in a columnar fashion
- Most OLTP databases will either be augmented by an in-memory database (IMDB)
- Most large-scale database servers will achieve horizontal scalability through clustering

*"If you're bringing back all the columns, a column-store database isn't going to perform any better than a row-store DBMS, but analytic applications are typically looking at all rows and only a few columns. When you put that type of application on a column-store DBMS, **it outperforms anything that doesn't take a column-store approach.**"*

- Donald Feinberg, Gartner Group

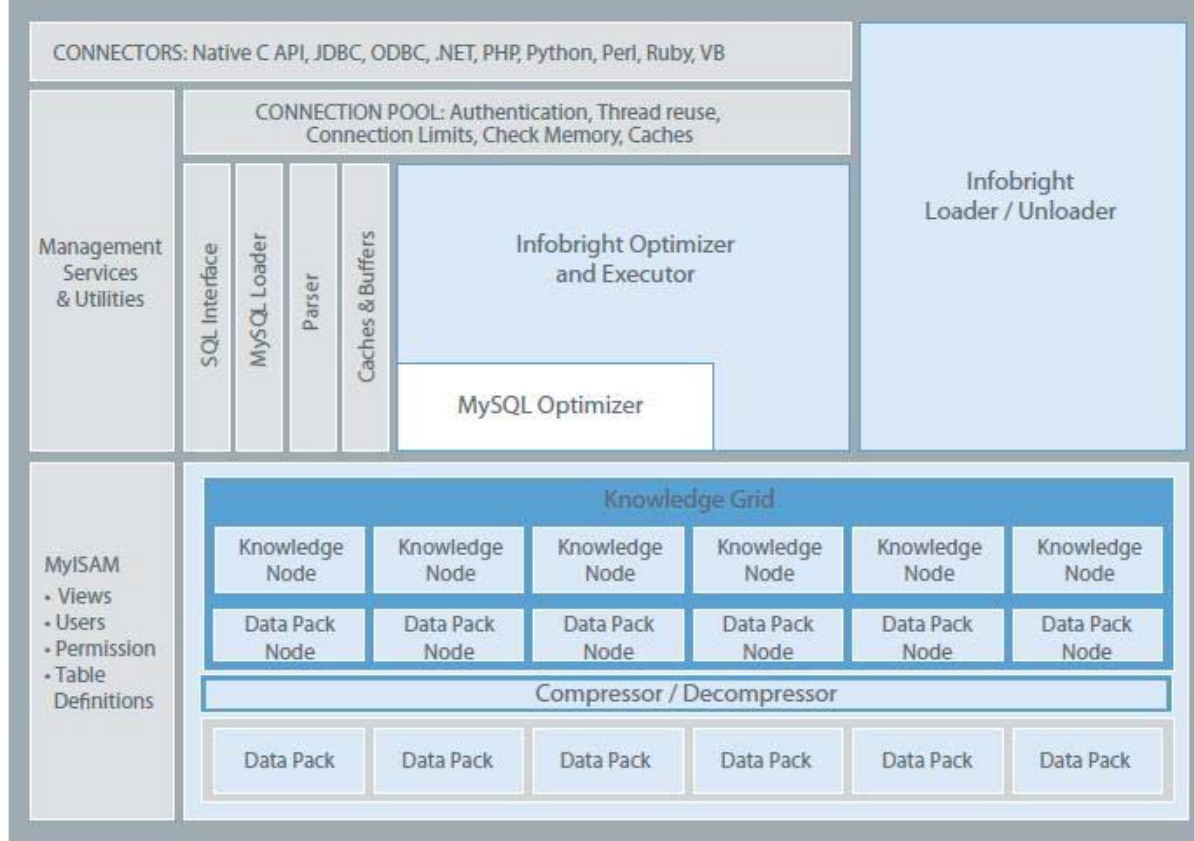
The Gartner logo, consisting of the word 'Gartner' in white text on a blue square background with a grid pattern.

Gartner

Infobright



Infobright Architecture



- Connectors
 - Standard MySQL
- Knowledge Grid
 - Node
 - Data Pack



Infobright

 Infobright.org



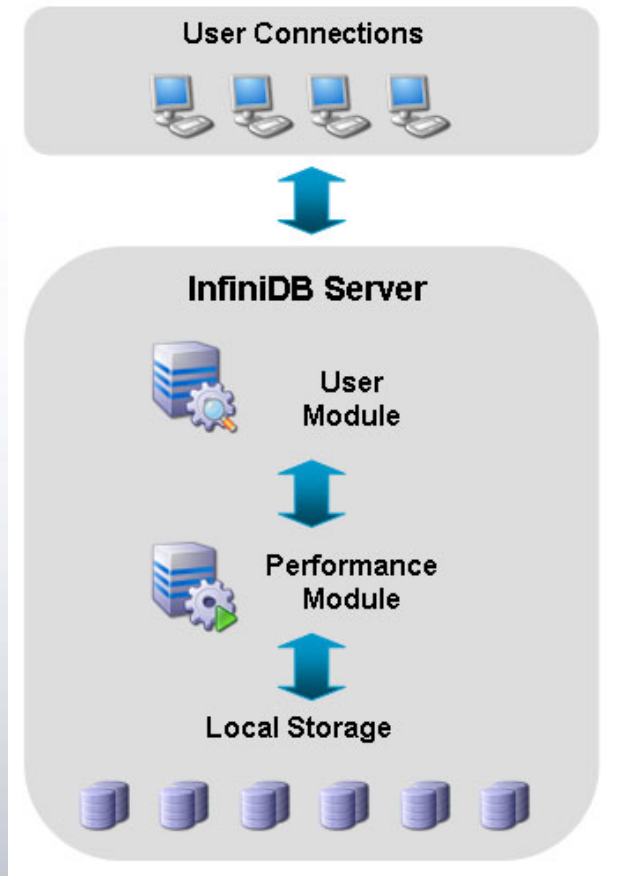
▪ Prednosti

- visoki stupanj kompresije
- poboljšana funkcionalnost rada SQL upita
- poboljšane performanse SQL upita
- kvalitetan meta-data sloj i arhitektura

▪ Nedostaci

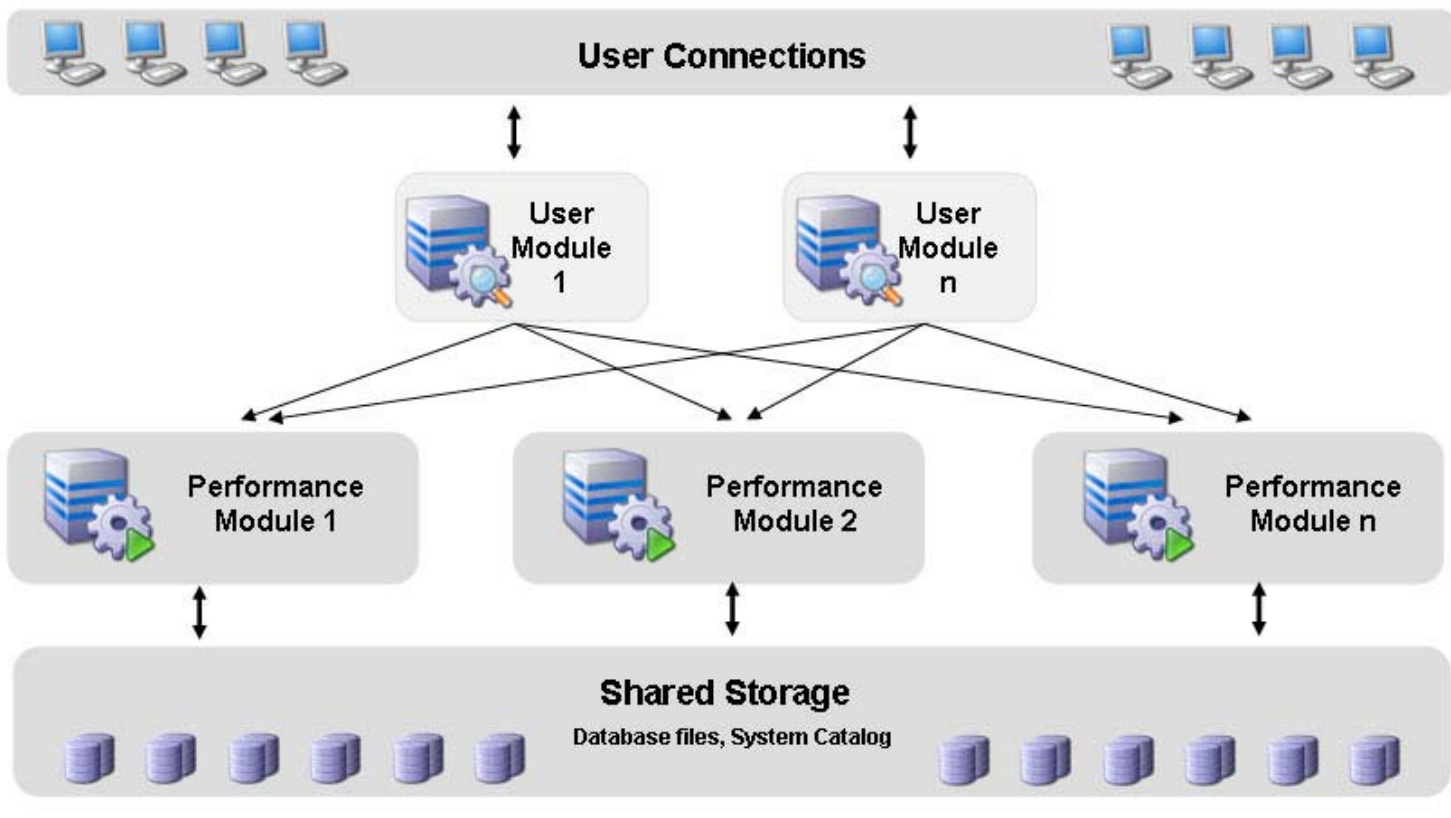
- koristi samo jedan procesor, nema MPP, loša skalabilnost
- community verzija ne podržava DML funkcije

InfiniDB – community version



- User Module
 - standard MySQL kompatibilnost
 - query procesiranje
- Performance Module
 - load i write procesiranje
 - data cache procesiranje
- Storage
 - blok, extent, segment, particija
 - logička kompresija podataka

InfiniDB – enterprise version



SKALABILNOST
PERFORMANCE

InfiniDB



▪ Prednosti

- odličan skup mogućnosti i visoke performanse
- skalabilnost na više procesora SMP (vertikalna) i MPP (vodoravna), ali ne u community verziji
- unaprijeđen rad optimizatora upita u odnosu na standardni MySQL

▪ Nedostaci

- ograničena kompatibilnost sa SQL standardima
- nema kompresiju podataka u fizičkom sloju

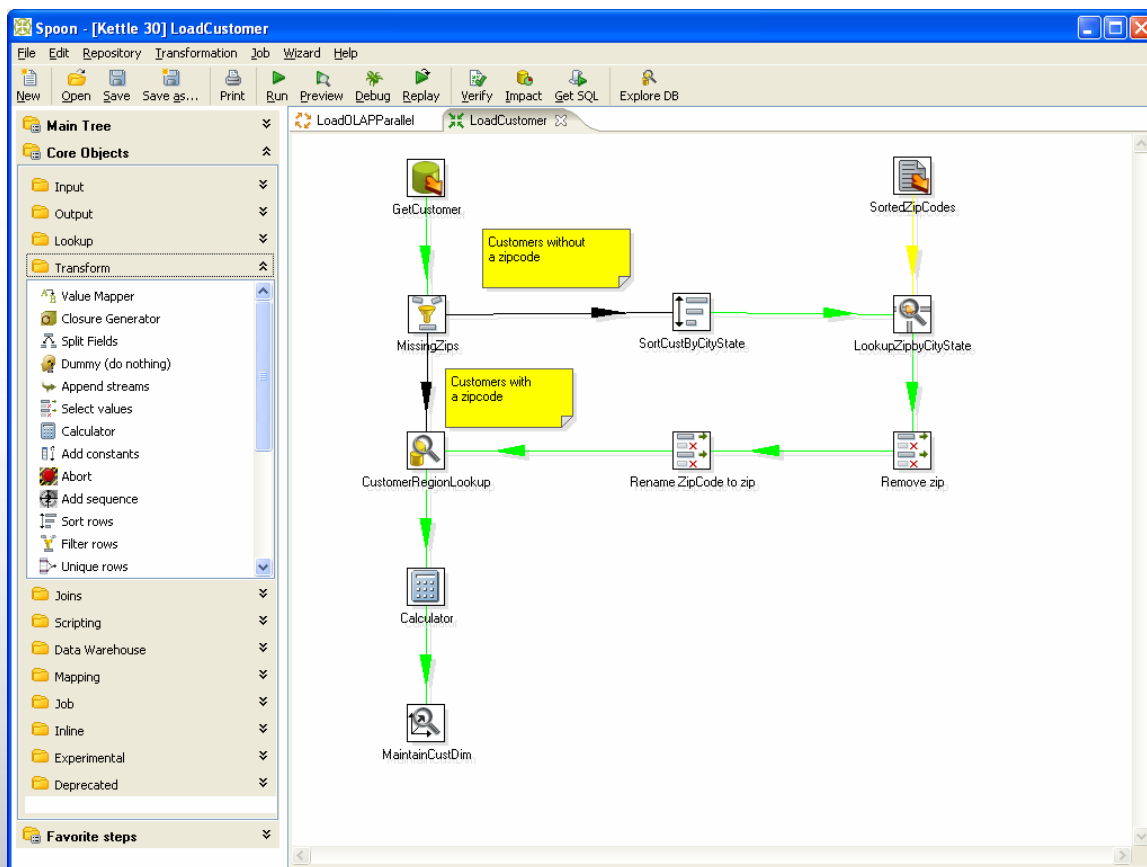
BI platforme s kolumnarnim bazama

OD15
HR OUG



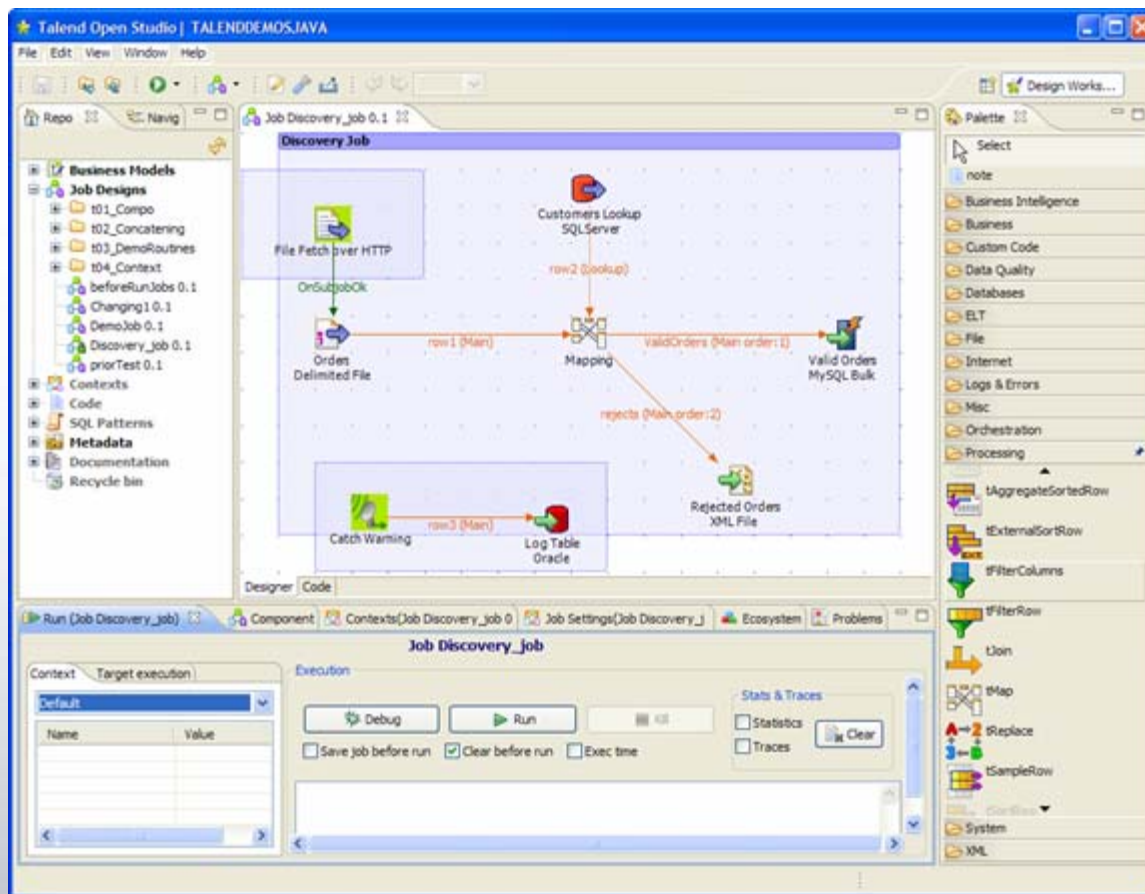
- Podrška za standardne protokole MySQL baze
 - osnova za podršku velikog broja BI alata
- Suradnja s vodećim OpenSource BI vendorima
 - Pentaho
 - Jaspersoft
- Alati za integraciju podataka
 - Pentaho kettle
 - Talend

Integracija podataka – Pentaho Kettle



INTEGRIRANO U PLATFORMU
MODULARNO

Integracija podataka - Talend



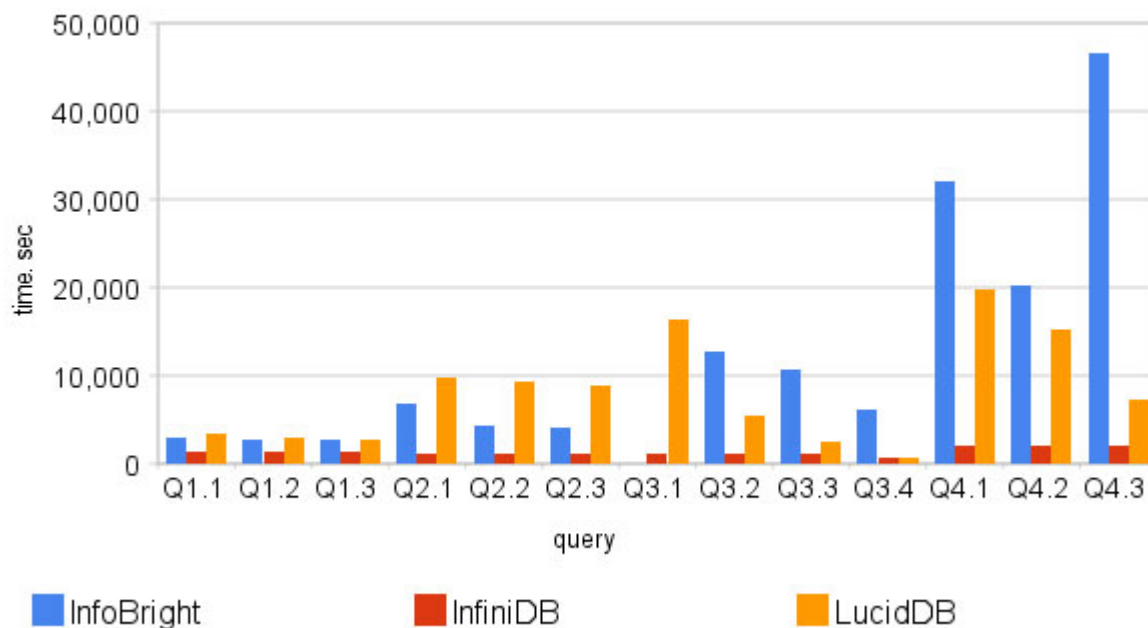
OPEN STUDIO
INTEGRIRANI DEBUGER
Q/A PODATAKA

Performanse



- Usporedba kolumnarnih baza za standardni skup benchmark testova

SSB queries



ZAUZEĆE PROSTORA

Infobright 112G

InfiniDB 626G

LucidDB 127G+indx

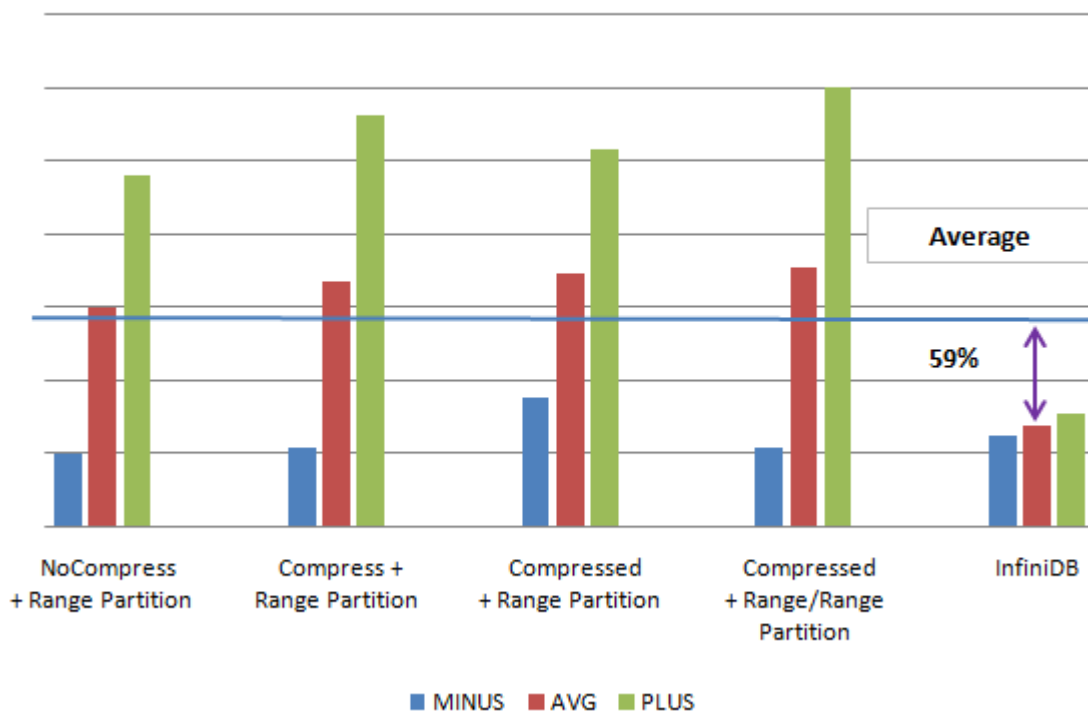
MonetDB 650G

InfiniDB performanse



Query Run Time Ranges

4 CPU, 8 GB RAM, 14 SATA HD 7,200 RPM



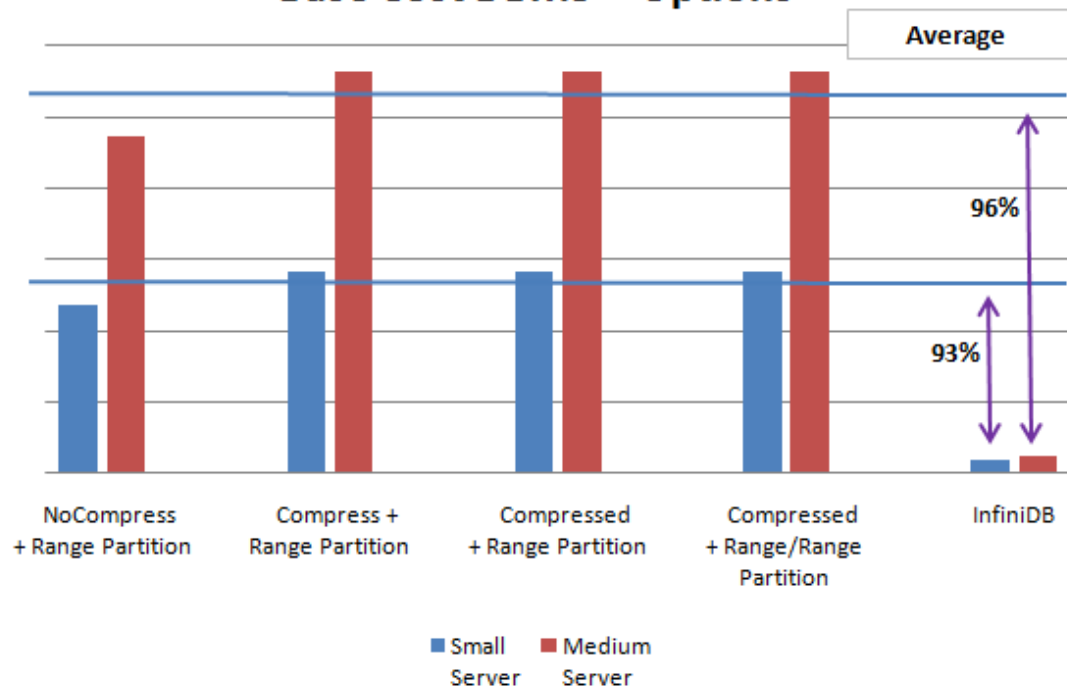
- 59% Brže izvršenje upita
- Nepotrebno dodavati i reorganizirati indekse
- Predvidljivost trajanja izvršenja upita - 95% garantiran odziv

Bert Scalzo, Oracle ACE

InfiniDB - cijena



Base Cost DBMS + Options



- TCO < 1/10 KDBMS
- Duplo brži odziv i deset puta povoljnije
- Implementacija jednaka kao MySQL – isti konektori – standard
- Manje održavanja - nema indeksa, nema particioniranja, nema predkalkuliranih vrijednosti

Zaključak

- Open Source DBMS
- Kolumnarna arhitektura
- Baza podataka namijenjena za DW
- Open Source integracija podataka
- 64bit hardver i softver
- Napredne analitičke mogućnosti
- Podrška i alati za Q/A podataka

