

NOSQL DATUBĀZES NOSQL DATABASES

Autors: **Jānis KRAPANS**, e-pasts: janis.kk1996@gmail.com
Studiju darba vadītājs: **Sergejs KODORS, Dr.sc.ing.**, e-pasts: sergejs.kodors@rta.lv
Rēzeknes Tehnoloģiju akadēmija
Atbrīvošanas aleja 155, Rēzekne, Latvija

***Anotācija:** Darbā ir aprakstīts trīs NoSQL datubāzu salīdzinājums: Mango, CouchBase un Cassandra. Datubāzu salīdzinājuma dati tika iegūti izmantojot YCSB- programmu, kura noslogo datubāzi ar dažādām darba slodzēm, kuras atbilst mūsdienu moderno aplikāciju prasībām. Testu izpildes beigās tiks iegūti katras datubāzes veikspējas dati, kā arī tās latentums.*

***Atslēgvārdi:** relāciju datubāzes, SQL, NoSQL datubāzes, efektivitāte, veikspēja.*

Ievads

Relāciju datubāzes ir vadošais modelis datu uzglabāšanai un pārvaldībai, piedāvājot savas ACID īpašības, bet tehnoloģijām attīstoties, prasības datubāzēm, kā mērogojamība, veikspēja palielinās, rezultātā ir uzsākta izstrāde citām alternatīvām- NoSQL datubāzes

Mūsdienās viss tiek glabāts datubāzēs. Īpaši izplatītas ir relāciju datubāzes un problēma mūsdienās ar ko saskarās relāciju datubāzes ir mērogojamība, tās ir izstrādātas lai strādātu uz viena servera lai saglabātu datu integritāti. Gadījumā ja ir sasniegti lietojamā servera limiti, lietotājam ir jāiegādājas dārgāka aparatūra ar lielāku apstrādes jaudu. No otras puses NoSQL datubāzes ir izstrādātas arī priekš darba uz vairākiem serveriem, kuri strādā kopā, sadalot slodzi starp tiem.

Darba mērķis: salīdzināt visbiežāk izmantojamas relāciju datubāzu alternatīvas.

Materiāli un metodes

Salīdzināšanas dati tika iegūti no avota internetā [1]. Tie tika realizēti ar *Yahoo Cloud Serving Benchmark (YCSB)* programmu [2], kura noslogo datubāzi ar dažāda veida darba slodzēm, lai pārbaudītu tās veikspēju. Salīdzinājumam tika izvēlēta trīs no populārākajām NoSQL datubāzēm:

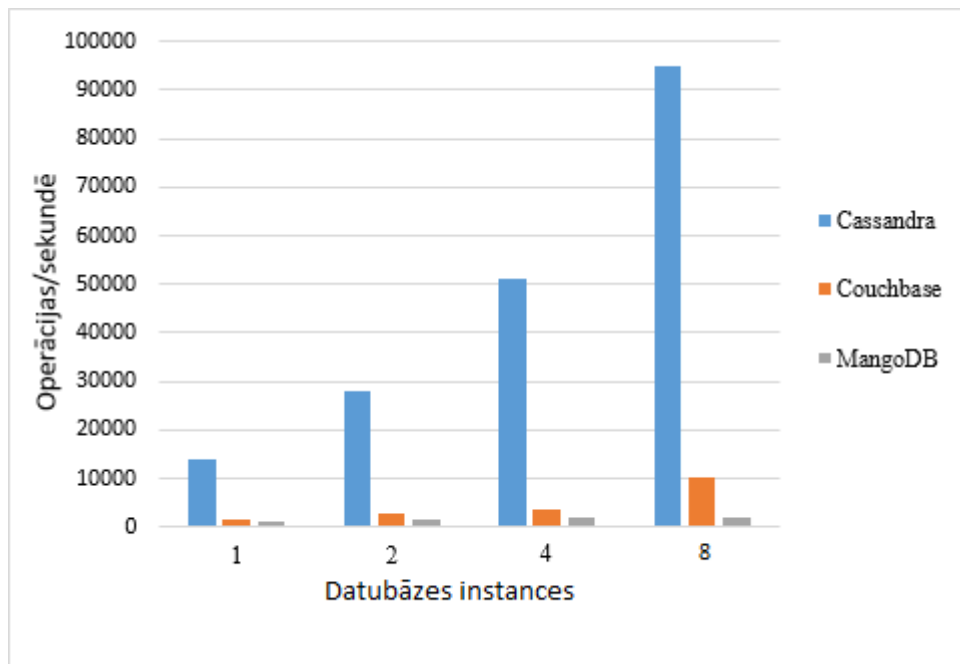
- Mango
- CouchBase
- Cassandra

Rezultāti

Datubāzes tika salīdzinātas pēc to darbības ātruma- operācijas sekundē, pie dažādām darba slodzēm. Datubāzes veikspēja arī tika pārbaudīta, kad tās veic darbu uz vairākām instancēm vienlaikus, kā arī tiks ņemts vērā latentums. Tika izmantoti 3 darba slodzes veidi, kuri atbilst mūsdienu moderno aplikāciju prasībām:

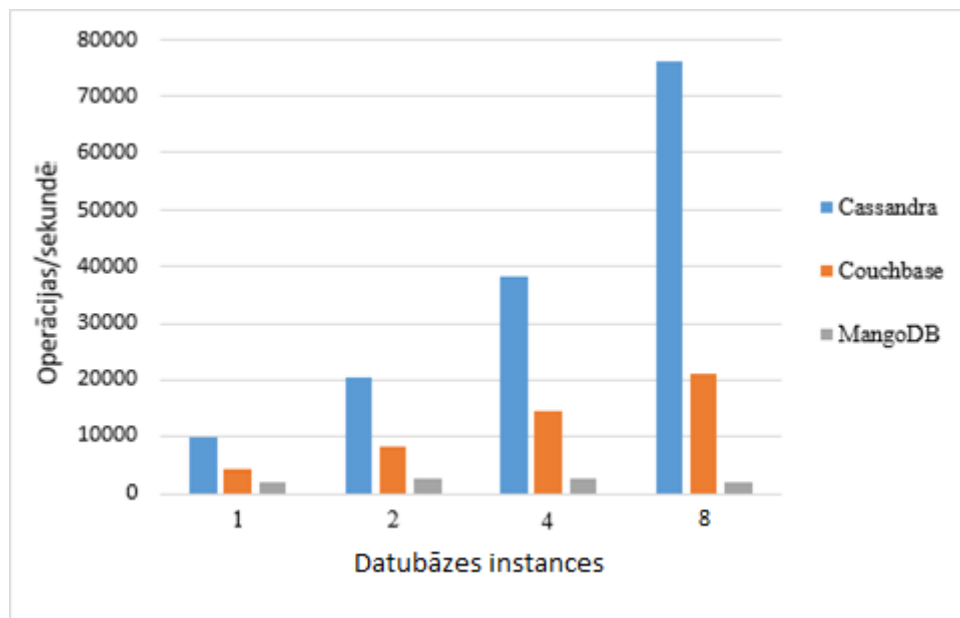
- Smagās atjaunināšanas noslodze
- Galvenokārt lasīšanas noslodze
- Lasīšana – modificēšana - rakstīšana

Smagās atjaunināšanas noslodzes testa rezultātus var apskatīt 1. attēlā.



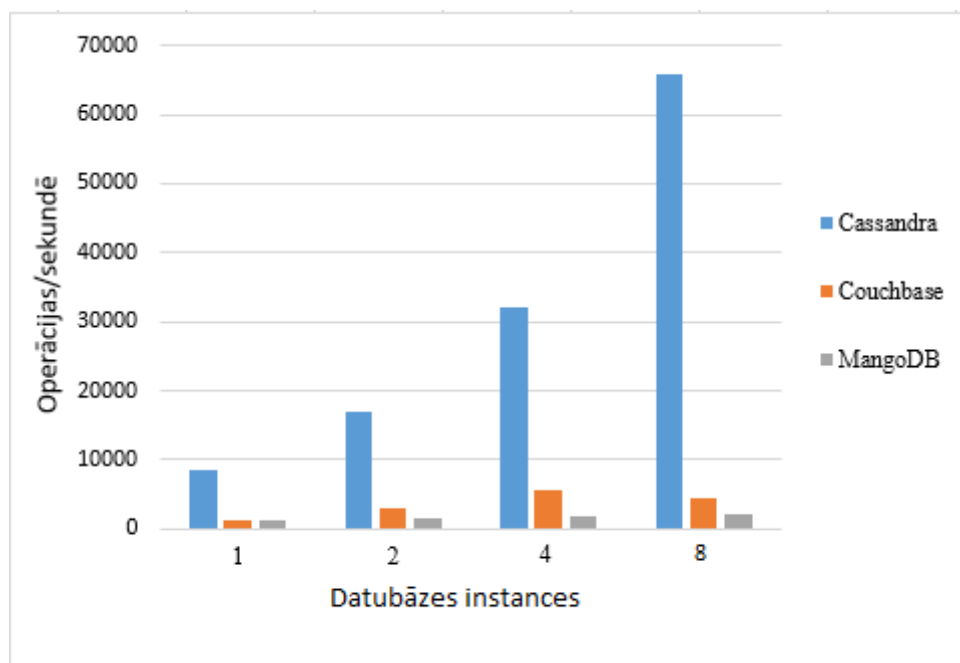
1. attēls. Smagās atjaunināšanas noslodzes rezultāti

Galvenokārt lasīšanas noslodzes testa rezultātus var apskatīt 2. attēlā



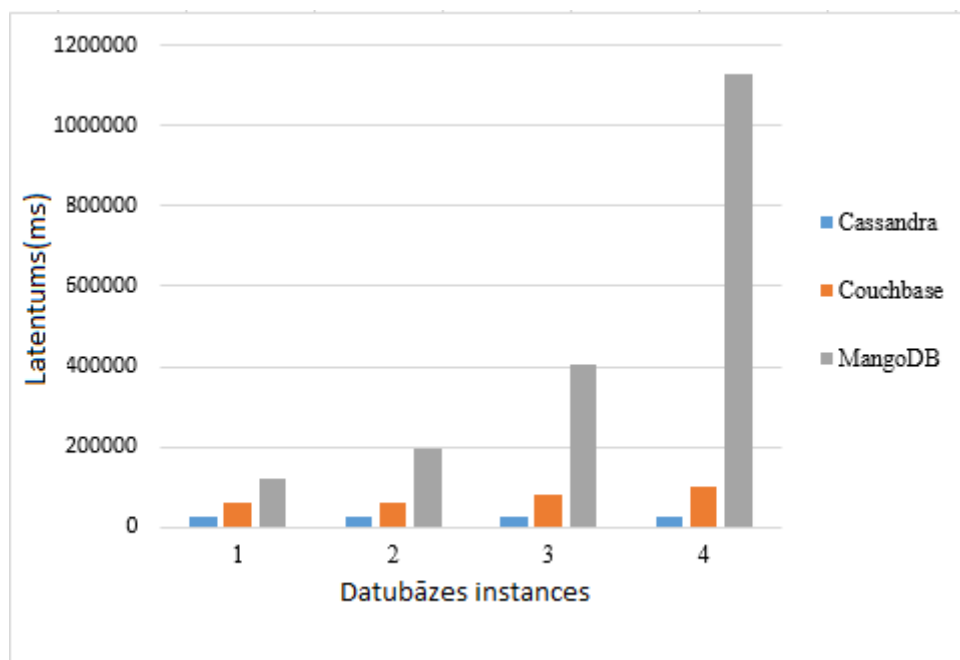
2. attēls. Galvenokārt lasīšanas noslodzes rezultāti

Lasīšanas – modificēšanas - rakstīšanas noslodzes testa rezultātus var apskatīt 3. attēlā



3. attēls. Lasīšanas – modificēšanas – rakstīšanas testa rezultāti

Datubāzu latentumu var apskatīt 4. attēlā



4. attēls. Datubāzu latentuma salīdzinājums

Pēc testu veikšanas tika iegūti dati, kuri rāda, ka katrā no noslodzes testiem, dominējošā datubāze ir Cassandra, kura vairākas reizes pārspēj, gan CouchBase, gan Mango datubāzi pat desmitkārtīgi. Ir jāņem vērā, ka šie testi tika veikti 2013. gadā un tie ir novecojuši. Cassandra datubāze parādīja labi savas labākās īpašības- horizontālo mērogojamību un milzīgu datu apjomu efektīvu apstrādi. Palielinoties instanču skaitam Cassandra datubāzes veiktspēja palielinājās līdz pat 99%

Secinājumi

- Darba gaitā tika izpētītas un aprakstītas *NoSQL* un *SQL* tehnoloģijas un tika salīdzinātas 3 no populārākajām *NoSQL* datubāzēm. Salīdzināšanas dati tika ņemti no informācijas avota, kurā bija aprakstīti šo datubāzu salīdzināšanas rezultāti. Salīdzināšana tika veikta izmantojot speciāli izveidotu atvērtā koda specifiku programmu: *YCSB*.
- Lai salīdzinātu datubāzes tika izmantoti 3 slodzes veidi: smagās atjaunināšanas noslodze, galvenokārt lasīšanas noslodze un lasīšana-modificēšana-rakstīšana
- Palielinoties instanču skaitam datubāzu veiktspēja palielinājās pat līdz 99%, bet testu rezultātos tika novēroti daži izņēmuma gadījumi, kad instanču skaitam palielinoties datubāzes efektivitāte samazinājās. Skatoties uz Smagās atjaunināšanas noslodzes testa rezultātiem redzams, ka *CouchBase* datubāzes veiktspēja uz 8 instancēm samazinās par aptuveni 18% no iepriekšējās veiktspējas testa rezultātiem uz 4 instancēm. Vēl viens izņēmums ir novērojams Galvenokārt lasīšanas noslodzes testa rezultātos, redzams, ka *MangoDB* datubāzes veiktspēja uz 8 instancēm samazinās par aptuveni 21% no iepriekšējās veiktspējas testa rezultātiem uz 4 instancēm.
- Vismazāko apstrādes laiku sasniedza *CassandraDB*, ar 26312,23 μs lielu aiztures laiku, kas ir aptuveni 2,6 milisekundes, datubāze ar vislielāko apstrādes laiku ir *MangoDB*-1540875,22 μs lielu aiztures laiku, kas ir aptuveni 154 milisekundes.

Bibliography

1. http://www.datastax.com/wp-content/themes/datastax-2014-08/files/NoSQL_Benchmarks_EndPoint.pdf skatīts internetā 02.06.2018
2. <https://github.com/brianfrankcooper/YCSB/wiki/Core-Workloads> skatīts internetā 02.06.2018