

IMPLEMENTIMI I NOSQL NË RRITJEN E PERFORMANCËS SË SHËRBIMEVE TË BIZNESEVE

SILVANA GRECA., ANXHELA KOSTA., SUELA MAXHELAKU.

Universiteti i Tiranës, Fakulteti i Shkencave të Natyrës, Departamenti i Informatikës

e-mail: silvana.greca@fshn.edu.al

Përmbledhje

Në ditët e sotme, me zhvillimin e teknologjisë, informacionet dhe të gjitha të dhënat me të cilat duhet të punojmë janë duke u shtuar me sasi të mëdha në një mënyrë të pabesueshme. Biznesi është ai që po përballet me rritjen e të dhënave dhe nevojën për të pasur sisteme shumë efikase për menaxhimin e tyre. Nga ana tjetër është edhe teknologjia ajo që ndihmon biznesin të përpunojë dhe të ruajë të gjitha sasi të mëdha të të dhënave në platforma të reja dhe në një mënyrë shumë efektive. Një kompani, për të sjellë risi dhe rezultate më të mira në biznes, ka nevojë për një menaxhim efikas dhe disponueshmëri të lartë të të dhënave edhe pse në sisteme të shpërndara. Teknologjia krijoi bazën e të dhënave NoSql për të na ndihmuar në menaxhimin e të dhënave të mëdha. Në këtë punim do të tregojmë një pasqyrë të përgjithshme të NoSql duke përfshirë përfitimet dhe sfidat e saj. Do të paraqesim se si bazat e të dhënave NoSql i mundësojnë bizneseve të ruajnë sasi të mëdha të dhënash në sisteme të shpërndara në mënyrë të shkëlqyer. Nëpërmjet një implementimi do të tregohet se si bazat e të dhënave NoSql ndihmojnë në rritjen e performancës së shërbimeve të bizneseve, duke ndërtuar një rast krahasues ndërmjet bazave të të dhënave relacionale dhe jo-relacionale.

Fjalëkyçe: NoSql, biznes, menaxhim i të dhënave, implementim, sisteme të shpërndara.

Abstract

Nowadays, with the development of the technology, the information and all the data that we need to deal with is becoming huge in an unbelievable way. The business is the ones which are facing the increasing of their data and the need to have very useful management systems to save them. From the other hand is also the technology which helps the business to process and store all the large amount of data in new platforms and in a very effective way. In order to bring innovation and great results, each company need an efficient data management system and high availability of them, even though in distributed systems. The technology created NoSql database to help us manage the large data. In this paper will be show a general overview of NoSql including its benefits and challenges. We will present how NoSql databases enable business to store large amounts of data in distributed systems and in a great manner. Through an implementation, it will be shown how NoSql databases helps to increase the performance of business services by demonstrated a comparative case between relational and non-relational databases.

Key words: NoSql, business, data management, implementation, distributed systems.

Hyrje

Në ditët e sotme, domosdoshmëria për web-in dhe zhvillimi i teknologjive të reja kanë bërë që sasi tepër të mëdha të dhënash të shtohen në sisteme të shpërndara. Rritja e të dhënave të pastrukturuara dhe nevoja për të pasur një menaxhim sa më efikas të tyre është një sfidë më vete për shumë biznese. Të gjithë llojet e bizneseve të industrive të ndryshme, jo vetëm bizneset e orientuara në web, janë shtyrë drejt të dhënave të mëdha, për tu përshtatur me teknologjitë e reja. Aplikacionet moderne që krijojnë dhe shfrytëzojnë të dhëna të mëdha në kohë reale, kanë sjellë një sfidë dhe transformim në menaxhimin e tyre, duke i kaluar kufijtë e sistemit të menaxhimit të bazës së të dhënave relacionale. Për të na ndihmuar në menaxhimin e të dhënave të mëdha do të përdoret një tip i ri i bazës së të dhënave, NoSql. Termi NoSql u përdor fillimisht nga Carlo Stozzi (1998) si emër i bazës së të dhënave të pastrukturuara. Në përgjithësi, studiuesit dhe zhvilluesit kuptuan krijimin e kësaj metodologjie për të synuar mbështetjen e shkallëzueshmërisë, performancës dhe qëndrueshmërisë, si pika kryesore të nevojshme nga aplikacionet e internetit të gjeneratës së ardhshme. Sot, bizneset në çdo industri, nga interneti tek shitjet me pakicë e deri tek shërbimet, po shfrytëzojnë teknologjinë e bazës së të dhënave NoSql për zhvillim më të shpejtë, kosto operacionale të reduktuara dhe veprime të shkallëzueshme.

Bazat e të dhënave NoSql

Shumë organizata përballen çdo ditë me sasi të mëdha të dhënash të pastrukturuara, si: dokumente, e-mail-e, multimedia dhe media sociale. Sistemet tradicionale e kanë gjetur të vështirë për tu kujdesur për këto pjesë të mëdha të të dhënave të pastrukturuara, pasi infrastruktura që ata kanë bëhet shumë komplekse për ruajtjen e këtyre të dhënave. Sipas Noel (Noel et al., 2016), me rritjen e numrit të biznese dhe të shërbimeve që ata ofrojnë filloi shtimi i sasive të mëdha të të dhënave. Krijimi dhe zhvillimi intensiv i aplikacioneve në internet, platformat si e-shopping ose portalet sociale, kërkojnë përpunim të shpeshtë e të shpejtë të sasive të mëdha të të dhënave për miliona përdorues në kohë pothuajse reale. Menaxhimi ideal i këtyre të dhënave është përdorimi i bazave të të dhënave jo relacionale duke çuar në zhvillimin e modeleve krejtësisht të reja të bazave të të dhënave. Këto modele janë quajtur si NoSql. Ishte Carlo Stozzi i cili në vitin 1998 përdori termin NoSql si një bazë të dhënash jo relacionale. Qëllimi i kryesor i Stozzi-t ishte ruajtja e sasive të mëdha të të dhënave (të ashtuquajtura big data) në një sistem të shpërndarë, jo relacional për të siguruar aksesim të shpejtë, performancë të lartë dhe tolerancë ndaj gabimeve (Zollman, 2012).

Sistemet e menaxhimit të bazës së të dhënave NoSql janë të dobishme kur punojnë me një sasi të madhe të të dhënash dhe kur natyra e të dhënave nuk kërkon një model relacional. Janë sisteme të shpërndara të dizenuara për ruajtjen e të dhënave në volum të lartë dhe për përpunimin masiv paralel të tyre në shumë servera. Ata përdorin gjuhët dhe mekanizmat jo Sql për të bashkëvepruar me të dhënat. Nga zhvillimi i aplikacioneve në web 2.0 të

cilat ishin projektuar për tu përdorur nga miliona përdorues, e pati motivin kryesor krijimi i NoSql për të kryer përditësime dhe lexime në kohë reale, në ndryshim nga bazat e të dhënave relacionale (Jatana et al., 2012)

Rritja e bizneseve nëpërmjet internetit dhe nevoja për NoSql

Çdo biznes që dëshiron të konkurojë me sukses në ekonominë e tregut duhet të plotësojë kërkesat e konsumatorit. Sigurimi i një konsumatori të kënaqshëm është një domosdoshmëri për çdo biznes. Interneti ka ndryshuar mënyrën se si funksionojnë bizneset dhe mënyrën se si individët kryejnë transaksionet e tyre. Aftësia për të komunikuar dhe për ta transferuar informacionin kudo në botë ka bërë që të krijohen aplikacionet e-commerce, të cilat janë në funksion të plotë të biznesit. Mënyra se si bizneset e vogla dhe të mesme shkëmbejnë informacione ka ndryshuar në mënyrë drastike pasi interneti është bërë popullor. Interneti ka bërë që të krijohen dhe të zhvillohen çdo ditë e më shumë teknologji të ndryshme që i vijnë në ndihmë botës së biznesit. Bizneset gjithmonë kanë si qëllim të arrijnë fitimet maksimale të tyre, për këtë arsye përpiqen të përshtaten dhe të update-ohen me teknologjitë e reja. Duke integruar këto teknologji, shtohen shumë informacionet dhe të dhënat (Dzhakishev, 2014).

Nga statistika të shumta që janë bërë (Vaish, 2013), është arritur në konkluzionin se me zhvillimin e internetit dhe futjen e cloud computing që të gjithë bizneset e llojeve të ndryshme të industrive kanë:

Të dhëna të mëdha që po rriten çdo ditë.

Nevojë për disponueshmëri të dhënash në kohë reale.

Nevojë për mbështetje të transaksioneve.

Nevojë për një model të të dhënave më fleksibël.

Nevojë për një arkitekturë më të mirë.

Meqenëse bazat e të dhënave relacionale nuk mundën të përballojnë këto të dhëna, lind nevoja për NoSql të cilat janë krijuar posaçërisht për të menaxhuar sasi të mëdha të dhënash në sisteme të shpërndara. Ndërsa Sql dhe NoSql mund të plotësojnë njëra-tjetrën në ekosistemin e sotëm të të dhënave të mëdha, shumica e bizneseve në industri të ndryshme duhen të përqendrohen në përdorimin e NoSql për të shpëtuar nga menaxhimi i sasive të mëdha të të dhënave (McCreary, 2014).

Përdorimi i NoSql në biznese

Një studim i kryer nga media kërkimore në treg (MRM, 2016) mbi rritjen e tregut të NoSql parashikon një normë të rritjes vjetore (CAGR) prej 21% për teknologjinë NoSql në vitet 2015-2018. Duke parë këtë statistikë mund të themi që NoSql është trendi i internetit. Të gjithë bizneset që kanë përdorur NoSql kanë treguar që NoSql është një model shumë fleksibël i të dhënave, ka shkallëzim horizontal dhe një arkitekturë të shpërndarë.

Një ndër përdorimet e NoSql është në kanalën e motit: "The Weather Channel" (TWC), për ti shërbyer miliarda shikuesve dhe përdoruesve në mbarë botën. Për të qenë sa më pranë klientit, duke mos sjellë vonesa në

kthimin e përgjigjes dhe duke mos ulur performancën, kanali i motit kërkonte një bazë të dhënash të shkallëzuar e transaksionale për të bërë menaxhimin e tyre. Përveç kësaj, disponueshmëria e vazhdueshme dhe aftësia për të kryer lexime dhe shkrime gjatë kryerjes së miliona transaksioneve në ditë nuk mund të garantohen me sistemet e bazave të të dhënave relacionale (Strickland, 2014).

Kanali i motit filloi të përdorte bazën e të dhënave NoSql, Apache Cassandra e cila e ndihmon të japë mesatarisht 100 milionë transaksione në ditë me një maksimum prej rreth 180-200 milionë transaksione ditore. Kanali i motit lejon përdoruesit ta ndajnë motin në kohë reale kudo që ndodhen dhe kjo falë funksionaliteteve të Cassandra-s (DataStax, 2014).

Numri i bizneseve dhe kompanive të mëdha që përdorin bazën e të dhënave NoSql është rritur. Disa prej tyre janë: Platformat e-commerce duke përfshirë Ebay dhe Wiggle të cilët e përdorin Cassandra-n për motorët e tyre të rekomandimit, NASA përdor NoSql për të ruajtur sasi të mëdha të të dhënave të sigurisë, GitHub dhe Instagram përdorin Cassandra për analizime, Barracuda dhe Iovation përdorin Cassandra për zbulimin e mashtrimit (Haughian, Osman and Knottenbelt, 2016).

Grupi Gartner (2012), krijoi hartën e bizneseve që përballen me volume të mëdha të dhënash si në figurën 1. Duke parë këtë hartë kuptojmë se sa mirë mund tu shërbejë bizneseve të ndryshme teknologjia NoSql duke pasur parasysh edhe karakteristikat kryesore që përcaktojnë Big Data-t siç janë: shpejtësia, ndryshueshmëria, volumi, kompleksiteti etj.

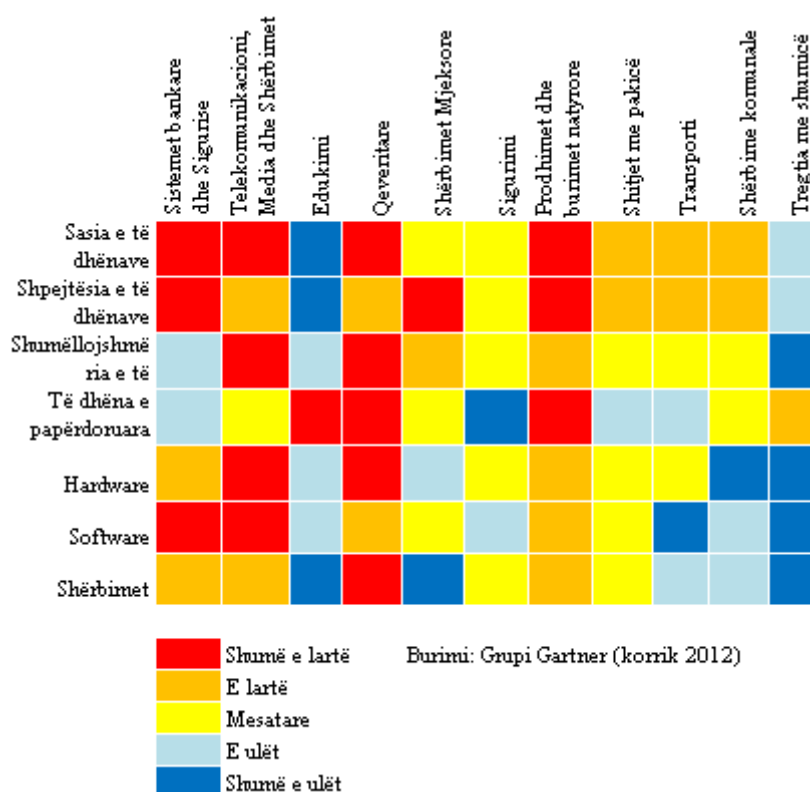


Figura 1. Harta e bizneseve që përballen me Big Data

Implementimi i NoSql në ndihmë të shërbimeve të bizneseve

NoSql është vërtetuar (Mongoddb, 2016) që është zgjidhja ideale e menaxhimit të të dhënave me vëllime të mëdha duke i mundësuar bizneseve përfitime në shërbime pasi i sjell atyre një performancë të lartë. Do të paraqesim një implementim për të kuptuar më mirë bazat e të dhënave NoSql i cili është realizuar në Apache Cassandra.

Cassandra është një sistem i shpërndarë për menaxhimin e të dhënave të pastrukturuara, në sasi shumë të mëdha, të cilat janë të shpërndara nëpër shumë servera. Cassandra u projektua për të përmbushur nevojat e problemit të kërkimit në Inbox nga Facebook. Duke qenë se sasia e të dhënave dhe shpejtësia e kërkimit të tyre në inbox ishte tepër e lartë, Facebook-u krijoi Cassandra-n për të rritur shpejtësinë e kërkimit në të. Arkitektura e Cassandra-s është një model i shpërndarjes peer-to-peer (Hewitt, 2010), pa asnjë pikë të vetme dështimi duke pasur kështu disponueshmëri të lartë dhe shkallëzimin horizontal. Të dhënat shpërndahen në mënyrë të barabartë në të gjithë nyjet për të garantuar balancimin e ngarkesës (Lakshman, Malik, 2010).

Në ditët e sotme bizneset e shohin shumë të arsyeshme dhe të nevojshme që klientët të ngelin të kënaqur kundrejt shërbimit që i ofrohet. Për këtë arsye janë ndërtuar aplikacione të ndryshme që i ofrojnë një shumëllojshmëri shërbimesh klientëve, në kohë reale. Këto aplikacione kanë nevojë për një mjedis, për të ruajtur të dhënat e mëdha që shtohen si shkak i ofrimit të shërbimeve. Kompleksiteti i lartë i të dhënave ka bërë që performanca kundrejt shërbimeve të ulet ndjeshëm duke shkaktuar vonesa dhe shpesh herë pritje të gjata të klientëve për marjen e shërbimit.

Implementimi i mëposhtëm tregon se si bazat e të dhënave NoSQL ndihmojnë në rritjen e performancës së shërbimeve të bizneseve duke bërë një menaxhim ideal të të dhënave edhe pse janë në sasi të mëdha, kundrejt bazave të të dhënave relacionale. Klienti shkon pranë “Kompanisë X” ose një institucioni qeveritar i cili ofron një shërbim, për shembull, telekomunikim, certifikatat e përbërjes familjare të viteve të ndryshme etj.. Klienti shpesh herë mund të presë në radhë për të marrë shërbimin që dëshiron dhe ka raste kur edhe nuk e mer shërbimin brenda ditës për arsye mbingarkese në sistem. “Pritja në radhë”, apo vonesa në shërbim janë çështje që prekin drejtpërsëdrejti çdo biznes apo organizatë fitimprurëse ose jo fitimprurëse si edhe institucionet e ndryshme qeveritare.

Një zgjidhje e duhur për këtë problem do të jetë implementimi i NoSQL, i Cassandra. Ndërtojmë Kyspace Shembull i cili ka si qëllim që të mbajë të gjitha të dhënat e ciklistëve. Për çdo ciklist do të ruhen të dhënat personale si edhe viti dhe ekipi ku bën pjesë ai në atë vit. Në modelin relational kjo paraqitet si një lidhje një me shumë.

Krijimi dhe përdorimi i kyspace-it:

```
cqlsh>CREATE KEYSPACE Shembull WITH replication = {'class':
'SimpleStrategy', 'replication_factor': 1};
```

```
cqlsh>Use Shembull;
```

Krijimi i tabelës paraqitet në figurën 2.

```
cqlsh:shembull> desc ciklisti;

CREATE TABLE shembull.ciklisti (
  id uuid PRIMARY KEY,
  ditelindja date,
  ekipet map<int, text>,
  emri text,
  gjatesia int,
  kombesia text,
  mbiemri text,
  pesha int
) WITH bloom_filter_fp_chance = 0.01
```

Figura 2. Tabela në Cassandra

Fusha ekipet i listuar në tabelë ka tip të dhënash map, ku vetë map ka një tip të dhënash integer që tregon vitin kur ciklisti është pjesë e një ekipi dhe një

tip të dhënash text që tregon emrin përkatës të ekipit. Të dhënat shtohen në tabelë duke përdorur komanda insert ose formatin Json.

```
cqlsh>INSERT INTO ciklisti (id, ditelindja, ekipet, emri, gjatesia, kombesia, mbiemri, peshja) VALUES (5b6962dd-3f90-4c93-8f61-eabfa4a803e2, '1990-02-20', {2017:'Partizani', 2016:'17 Nentori', 2015:'Teuta', 2014:'Vllaznia', 2013:'Flamurtari'}, 'Miri', 170, 'Shqiptare', 'Prifti', 65 );
```

Figura 3 ilustron një query në Cassandra.

```
cqlsh:shembull> select emri, ekipet from ciklisti;
emri | ekipet
-----|-----
Miri | {2013: 'Flamurtari', 2014: 'Vllaznia', 2015: 'Teuta', 2016: '17 Nentori', 2017: 'Partizani'}
```

Figura 3. Marrja e të dhënave në Cassandra

Pjesa më e madhe e shërbimeve që ofrohen nga biznesi ka të bëjë me kohën e pritjes. Koha e pritjes tregon edhe performancën e çdo biznesi.

Në shembullin tonë kemi gjeneruar disa të dhëna të mëdha random për mbushjen e tabelës për të parë kohën e ekzekutimit në Apache Cassandra.

Në MySQL kemi krijuar dy tabela, tabelën çiklisti, e cila mban të dhënat personale të çiklistëve dhe tabelën ekipet që ka id e çiklistit, vitin dhe emrin e ekipit përkatës në atë vit. Të njëjtat të dhëna si në Cassandra i kemi përdorur për të populluar tabelat edhe në MySQL në mënyrë që të shohim se cila prej tyre është më performante duke ekzekutuar query të ndryshme dhe në të njëjtën platformë harduerike.

Në rastin e NoSql, figura 4-a paraqet kohën e shkrimit të kërkesës, pra shtimi i të dhënave në tabelën përkatëse, ndërsa figura 4-b paraqet kohën e kthimit të përgjigjes kur kërkojmë të dhëna sipas specifikave të caktuara.

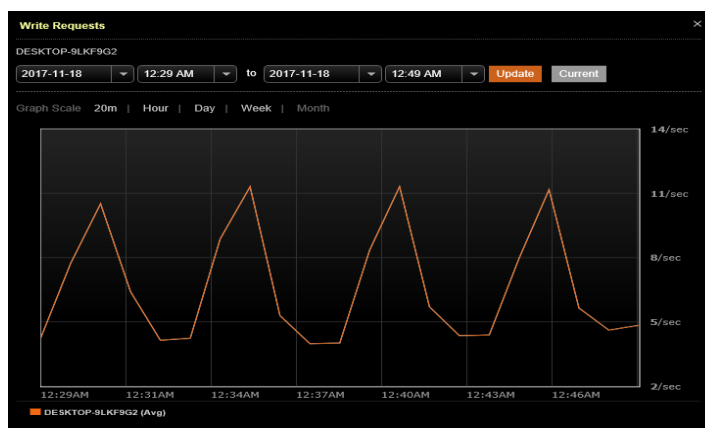


Figura 4-a. Koha e shkrimit të kërkesës



Figura 4-b. Koha e kthimit të përgjigjes

Në secilën prej bazave të të dhënave kemi ekzekutuar query-ve të ndryshme, veprime që realizohen në SQL dhe CQL, koha mesatare e ekzekutimit e të cilave paraqitet në tabelën 1. Vëmë re se që koha e përgjigjes është shumë më e shpejtë në rastin e Cassandra's se sa në rastin e MySQL.

	Insert	Update	Delete	Select
Apache Cassandra	0.01 s	0.01 s	0.01 s	0.0015 s
MySQL	0.1208 s	0.1231 s	0.0919 s	0.0019 s

Tabela 1. Krahasimi i kohës së ekzekutimit, Cassandra vs MySQL

Tabela 2 paraqet rezultatet e këtij krahasimi si edhe disa nga tiparet e rëndësishme që kanë bazat e të dhënave NoSql dhe Sql. Nga rezultatet e arritura themi se bizneset me Big Data kanë përfitime më të larta duke përdorur NoSql sesa bazat e të dhënave relacionale.

	Performanca	Disponueshmëria	Kosistenca	Ruajtja	Shkallëzimi
Apache Cassandra	E lartë	E mirë	E ulët	I optimizuar për Big data	I lartë
MySQL	E ulët	E mirë	E mirë	Madhësi mesatare	I lartë por i kushtueshëm

Tabela 2. Krahasimi i Cassandra's me MySQL

Përfundime

Bazat e të dhënave relacionale po bien ndjeshëm kundrejt menaxhimit të të dhënave me volum të lartë. Bizneset e kanë të nevojshme përdorimin e internetit dhe të teknologjive, për këtë arsye ata updatohen gjithmonë me teknologjitë e reja. Mirëpo, përshtatja me teknologjitë e reja bën të mundur që sasia e të dhënave të rritet me hapa shumë të larta. Shërbimet që bizneset ofrojnë, fillojnë dhe ngadalësohen ndjeshëm pasi është i vështirë menaxhimi i të dhënave të mëdha. Është krijuar NoSql për ti ardhur në ndihmë të gjithë bizneseve të industrive të ndryshme që përballen me volume të larta të

dhënash. NoSql kanë shumë avantazhe për sa i përket menaxhimit të të dhënave krahasuar me të metat e tyre.

Nëpërmjet implementimit është treguar se përdorimi i NoSql (Cassandra) ndihmon në rritjen e performancës së shërbimeve të bizneseve duke bërë një menaxhim efikas të të dhënave të mëdha në kohë reale dhe shumë të shpejtë në krahasim me bazat e të dhënave relacionale (MySQL). NoSql të lejon monitorim të drejtpërdrejtë të bazës së të dhënave duke parë të gjitha nyjet e saj. Madje në momentin e dështimit të një nyjeje puna fillon pa prishur lidhjen e nyjeve të tjera. NoSql konsiderohet si baza e të dhënave të internetit duke nënkuptuar që me rritjen e internetit rritet edhe NoSql. E ardhmja e NoSql pritet të jetë shumë sfiduese duke zënë një pjesë shumë të rëndësishme në menaxhimin e të dhënave në të gjithë botën.

Literatura

Jatana, N., Puri, S., Ahuja, M., Kathuria, I., Gosain, D., (2012): A Survey and Comparison of Relational and Non-Relational Database. *International Journal of Engineering Research & Technology (IJERT)*, vol. I, no. 6, 2012: 1-5

Zollman, J., (2012) : NOSQL Databases: 3-17 [Online]: http://sewiki.iai.unibonn.de/_media/teaching/labs/xp/2012b/seminar/10-nosql.pdf

Lakshman, A., Malik, P. (2010): Cassandra - A Decentralized Structured Storage System. *ACM SIGOPS Operating Systems Review*, v.44 n.2, April 2010:1-6

Hewitt, E. (2010): Cassandra: The definitive guide. O'Reilly Media 2010: 223-234

BĂZĂR, C., Sebastian, C. (2014): The transition from rdbms to nosql. a comparative analysis of three popular non-relational solutions: Cassandra, mongodb and couchbase. *Database Syst J.* 2014; 5(2):49–59.

Priyanka, AmitPal (2016): A Review of NoSQL Databases, Types and Comparison with Relational Database. *International Journal of Engineering Science*, Volume 6 Issue No. 5, 2016: 4963-4966

Venkatraman, S., Fahd, K. and Venkatraman, R.(2016): SQL Versus NoSQL Movement with Big Data Analytics. *International Journal of Information Technology and Computer Science (IJITCS)*, vol.8, no.12, 2016:59-66

Dzhakishev, D (2014): NoSQL Databases in the Enterprise.:54-74

Haughian, G., Osman, R. and Knottenbelt, W. J. (2016): Benchmarking Replication in Cassandra and MongoDB NoSQL Datastore. *International Conference on Database and Expert Systems Applications*, 2016: 152-166.

Pokorny, J., (2013): NoSQL databases: a step to database scalability in web environment. *International Journal of Web Information Systems* Vol. 9 No. 1, 2013: 69-82

Moniruzzaman A.B.M., Hossain A.S., (2013): Nosql database: New Era of Databases for Big data Analytics- Classification, Characteristics and Comparison. *International Journal of Database Theory and Application*. Vol. 6, no. 4, 2013:1-14

Nati, Sh. (2009): The Common Principles Behind The NOSQL Alternatives.

Yuhanna, N (2016): The Forrester Wave™: Big Data NoSql, Q3 2016. Scalability, Multimodal, Flexibility, And low cost Drive Enterprice Adoption Of NoSql. Forrester Research 2016: 2-17

Vaish, G. (2013): Getting Started with NoSql, Your guide to the world an the technology of NoSql. Packt Publishing, 2013: 51-80

Mongodb, (2016): MongoDB Enterprise Advanced Datasheet, 2016: 1-2

Market Research Media, (2016): NoSQL Market Forecast 2015-2020