

Studi Perbandingan Performansi Antara MongoDB dan MySQL Dalam Lingkungan Big Data

Apri Junaidi
Fakultas Teknik
Prodi Teknik Informatika Univeristas Widyatama
Jl. Cikutra No. 204, Bandung, Indonesia
apri.junaidi@widyatama.ac.id

Abstract—Tiada hari tanpa, penggunaan data yang terus berkembang diikuti dengan kemajuan teknologi dibidang dalam teknologi informasi, sehingga kebutuhan untuk pengolahan data sangat diperlukan, proses pertambahan data yang terus meningkat bersumber dari berbagai macam transaksi seperti data posting blog, data percakapan di jaringan selular, sms, chat dan banyak sumber data lain yang setiap saat terus bertambah, sehingga terkenal dengan nama Big Data. Dalam pengolahan big data ini diketahui adanya database NoSQL, sedangkan database yang selama ini banyak dipakai adalah MySQL. Untuk mengetahui performansi response time masing-masing database maka diperlukan sebuah penelitian untuk mendapatkan hasil performansi tersebut. Penelitian ini menggunakan database dengan jumlah record yang besar yaitu sebanyak 226.232 record. Untuk keluarga NoSQL penulis menggunakan MongoDB dan untuk keluarga SQL penulis menggunakan MySQL, pengujian dilakukan pada lingkungan yang seimbang, pada hardware dan software yang memiliki spesifikasi yang sama. Hasil akhir penelitian menunjukkan bahwa MongoDB menunjukkan performansi yang baik dibandingkan MySQL.

Keywords—*Big Data; NoSQL; SQL; MongoDB; perbandingan studi;*

I. PENDAHULUAN

Pertambahan data yang terus meningkat bersumber dari berbagai macam transaksi seperti data posting blog, data percakapan di jaringan selular, sms, chat dan banyak sumber data lain yang setiap saat terus bertambah, sehingga terkenal dengan nama Big Data. Dalam pengolahan big data ini diketahui adanya database NoSQL, sedangkan database yang selama ini banyak dipakai adalah MySQL. Untuk mengetahui performansi response time masing-masing database maka diperlukan sebuah penelitian untuk mendapatkan hasil performansi tersebut. Penelitian ini menggunakan database dengan jumlah record yang besar yaitu sebanyak 226.232 diharapkan dengan jumlah data yang besar mampu memberikan performansi yang baik.

Mengenali dan memahami bagaimana menggunakan format data standar adalah penting untuk menjadi seorang ilmuwan data. pengolahan data dasar dan format file data yang berbeda,

seperti fixed-width teks, teks delimited, XML, JSON, CSV dan HDF[1]. Sebagai contoh praktis, penulis mencoba untuk memberdayakan sumber data ini untuk import kedalam format yang diperlukan dalam penelitian ini.

Big data merupakan daerah penelitian belum dijelajahi yang telah membawa tantangan baru dan menarik. Tantangan baru fokus terutama pada masalah yang berkaitan dengan bisnis dan teknologi [2][3].

Tiada ada hari tanpa data, Oleh karena itu, disiplin database yang membentuk salah satu pilar ilmu tentang data. Sistem Manajemen Database melayani dua tujuan: (1) Data penataan sehingga mereka dapat dikelola dengan mudah dan (2) menyediakan skalabilitas dan kinerja yang handal sehingga tidak perlu khawatir tentang penyimpanan data. Sampai saat ini, database relasional dan Structured Query Language (SQL). Karena pertumbuhan volume data, secara besar-besaran didistribusikan basis data- dan disebut database NoSQL [4].

Dalam penelitian ini, penulis membagi makalah ini dalam beberapa bagian: bagian pertama pendahuluan, yang berisi tentang pengantar yang berhubungan dengan penelitian ini, bagian kedua landasan teori yang memaparkan tentang teori pendukung penelitian ini, bagian ketiga tentang penelitian terdahulu yaitu pembahasan beberapa penelitian yang memiliki kesamaan dengan makalah ini, bagian keempat tentang studi perbandingan yang menjelaskan tatacara melakukan penelitian dalam makalah ini, bagian kelima menjelaskan tentang hasil penelitian yang dilengkapi dengan data capaian serta grafik pendukung, bagian keenam penutup, yang menjelaskan tentang kesimpulan dari penelitian dan saran yang dapat dilakukan pada penelitian berikutnya.

II. LANDASAN TEORI

2.13. *Big Data*

Kita hidup di Era Big Data. Dalam beberapa tahun terakhir, jumlah total data yang dibuat oleh manusia telah meledak. Dari tahun 2005 sampai 2020, jumlah data diprediksi akan meningkat 300 kali, dari 130 exabyte 40.000 exabytes [5]. Data adalah fakta dan statistik yang dikumpulkan bersama-sama untuk referensi atau analisis. Informasi adalah

Prosiding
ANNUAL RESEARCH SEMINAR 2016
6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

http://ars.ilkom.unsri.ac.id

fakta yang tersedia atau belajar tentang sesuatu atau seseorang. Keduanya dapat digunakan sebagai dasar untuk penalaran atau perhitungan. Big Data besar adalah volume tinggi (high-volume), tinggi-kecepatan (high-velocity) dan high-variety yang menuntut, bentuk-bentuk inovatif dalam pengolahan informasi hemat biaya, untuk ditingkatkan wawasan dan pengambilan keputusan [6].

Tantangan teknis untuk Big data termasuk persyaratan yang kuat real-time analisis, model penyimpanan baru, Data ini harus menjadi server yang dikelola atau cloud computing, komparasi dan divisualisasikan oleh para peneliti, hardware / software, ilmuwan komputer, dan statistik [7].

Dalam makalah ini salah satu tantangan big data dipaparkan yaitu, bagaimana melakukan perbandingan atau komparasi pada suatu data yang ada pada keluarga Big Data.

2.14. NoSQL

Istilah NoSQL diciptakan oleh Carlo Strozzi pada tahun 1998 dan mengacu pada database non-relasional, pada tahun 2009 Eric Evans memperkenalkan kembali istilah NoSQL. Baru-baru ini, istilah ini memiliki makna lain, yaitu "Not Only SQL", istilah yang lebih baik dari sebelumnya yang lebih dikenal dengan "anti-relasional"[8][9], NoSQL mengakomodasi tanda yang tidak terstruktur, kehadirannya bukan untuk menggantikan SQL namun kedua teknologi ini dapat saling berdapampingan. Perbedaan utama kedua database ini adalah SQL memiliki skema yang kaku sementara database NoSQL menawarkan desain yang fleksibel yang dapat diubah tanpa downtime atau gangguan layanan. NoSQL juga dirancang untuk menyimpan data yang didistribusikan untuk kebutuhan data dalam skala besar; misalnya Facebook memiliki 500 juta pengguna dan Twitter terakumulasi terabyte data, database NoSQL telah memiliki popularitas yang tinggi, sehingga database ini diklaim lebih baik dari database SQL [10][11], Database NoSQL dimotivasi oleh kesederhanaan, horizontal scaling dan kontrol yang lebih dalam kesediaan data [12].

NoSQL menjadi solusi dalam penanganan data dalam jumlah besar yang berkembang pesat saat ini. Data ini biasanya non-terstruktur, kompleks dan tidak cocok digunakan dalam model relasional. Contoh data yang bisa kita rasakan adalah data yang berasal dari smartphone yang mencatat lokasi broadcast setiap saat, video dan kamera bahkan halaman halaman website yang berisi banyak informasi serta dokumen[13].

2.15. Klasifikasi model data NoSQL

Klasifikasi model data NoSQL terdiri dari: Key Value Stores, Document Stores, Column Family Stores and Graph Database [8][14][15][16].

Model	NoSQL Database
Key Value Stores	Voldemort, Redis, Membase
Document Stores	Riak, MongoDB, CouchDB,
Column Family Stores	Cassandra, HBase, Hypertable
Graph Database	Sesame, BigData, Neo4J, GraphDb, FlockDB

- Key Value Stores- Key value stores memiliki kesamaan dengan pemetaan data atau direktori dimana data ditangani oleh sebuah kunci unik, dalam pelaksanaannya sangat sederhana yaitu berdasarkan atribut kunci saja.
- Document Stores- Document Stores merupakan data dalam bentuk dokumen JSON, lebih fleksibel. Secara konsep data ini memiliki jenis : JSON, BSON, XML and BLOBs.
- Column Family Store- Column Family Stores juga dikenal dengan data yang berorientasi kolom. Media penyimpanan ini terinspirasi oleh Googles Bigtable. Data disimpan dalam kelompok sel dalam sebuah kolom, dan kolom tersusun dari group kolom. Secara teori jumlahnya tidak terbatas.
- Graph Database – Database graph efisien dalam penyimpanan data dalam bentuk grafik atau gambar.

2.16. Fitur NoSQL

Keuntungan dari NoSQL adalah sebagai berikut: membaca dan menulis data dengan cepat, dukungan untuk penyimpanan besar, mudah untuk dikembangkan dan biaya rendah. Fitur dari NoSQL memiliki model data dan teori CAP yang diperkenalkan oleh Profesor Eric Brewer pada tahun 2000. Teori CAP singkatan dari Consistency, Availability, tolerance of network Partition.

2.17. MongoDB

MongoDB adalah sebuah database yang berorientasi dokumen lintas platform. MongoDB dikelompokkan kedalam NoSQL database, memberikan kemampuan untuk query pada bidang apapun dalam dokumen [17][18].

MongoDB is a cross-platform document-oriented database system. Classified as a "NoSQL" database, MongoDB bukan database relasional database, MongoDB support untuk dokumen JSON yang membuat integrasi data dalam beberapa jenis aplikasi lebih mudah dan lebih cepat [19].

MongoDB di release dibawah kombinasi lisensi GNU Affero General Public License dan, MongoDB free untuk digunakan dan opensource [11], ditulis dalam bahasa C++ dan disupport 10gen [20].

TABLE V. NoSQL DATA MODEL

Prosiding ANNUAL RESEARCH SEMINAR 2016

6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

Fitur dari MongoDB: Ad hoc queries, Indexing, Replication, Load balancing, File Storage and Aggregation [15].

2.18. MySQL

MySQL adalah sistem manajemen database relasional yang open source. Database ini banyak digunakan pada skala menengah sampai medium pada single server. MySQL lebih sederhana dan kemudahan penggunaan melalui software open source seperti phpMyAdmin.

Fitur dari MySQL: Cross platform support, Store procedures, Query Caching, Replication Support, Atomicity, Consistency, Isolation and Durability (ACID), Multiple Storage Engines [11].

2.19. JSON

JavaScript Object Notation (JSON). adalah format data-interchange yang ringan. Sangat mudah dibaca dan ditulis bagi manusia dan mudah juga mesin atau komputer untuk memarsing dan mengolahnya. JSON merupakan subset dari Bahasa Pemrograman JavaScript, Standard ECMA-262 3rd Edition - Desember 1999 [13][17].

2.20. CSV

Comma Separated Values (CSV) merupakan format data dalam basis data di mana setiap record dipisahkan dengan tanda koma (,) atau titik koma (;). data dalam format csv berbentuk sederhana sehingga data dalam format csv dapat dibuka dengan berbagai text-editor seperti Notepad, Wordpad, bahkan MS Excel juga mudah diimport kedalam database SQL dan NoSQL [21].

III. PENELITIAN TERDAHULU

Beberapa penelitian sebelumnya sudah dilakukan oleh para peneliti diantaranya:

Sebuah studi perbandingan MongoDB vs. MySQL [8], peneliti melakukan beberapa teknik perbandingan seperti query select, query insert, query update dan query delete pada kedua database menggunakan 5.000 record untuk setiap table untuk dilakukan pengujian, penelitian ini menggunakan sistem operasi Windows & Ultimate 64-bit, prosessor Intel Core i3 (2.4 Ghz), RAM 4GB. Hasil akhir menunjukkan bahwa MongoDB memberikan eksekusi waktu yang rendah dibandingkan MySQL pada kepada semua Query yang diberikan.

Comparing NoSQL MongoDB to an SQL DB [13], pada penelitian ini peneliti menginstal MongoDB dan Microsoft SQL Server Express pada komputer dengan prosessor Core i7 quad core 3.4 GHz, 8GB DDR3 RAM, database di install pada Solid State Drive (SSD) untuk mendapatkan kemungkinan

kecepatan baca dan tulis, untuk proses testing peneliti membangun sebuah aplikasi dalam bahasa C# menggunakan Visual Studio 2012 .NET Framework. Test terdiri dari 4 bagian test dengan 100 putaran untuk setiap test. Data terdiri dari 3 table/collection yang merupakan data department, project dan user. Hasil akhir menunjukkan MongoDB solusi yang baik untuk data dalam jumlah besar.

Performance Analysis for NoSQL and SQL [11], peneliti menggunakan perbandingan NoSQL dan SQL, peneliti menggunakan bahasa Java untuk membuat aplikasi perbandingan, database yang digunakan MongoDB, MySQL, VoltDB atau FoundationDB. Hasil akhir peneliti mendapatkan bahwa FoundationDB yang masih tergolong database NoSQL lebih baik dibandingkan SQL.

IV. STUDI PERBANDINGAN

Dalam studi ini penulis melakukan perbandingan untuk database NoSQL dan SQL, untuk NoSQL penulis menggunakan MongoDB, untuk keluarga SQL penulis menggunakan MySQL.

2.21. Dataset

Dataset yang digunakan untuk studi ini penulis dapatkan dari website *Crowdflower.com* yang bisa diakses di URL: <https://www.crowdflower.com/data-for-everyone/>

Dataset yang digunakan berisi tentang kumpulan data image, dataset terdiri dari 10 kolom dan 226.232 record, ukuran file pada saat didownload 55.5 MB. File asal dalam format CSV.

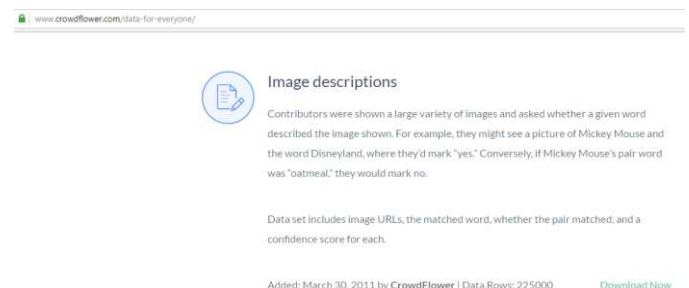


Fig. 9. Website Dataset CrowdFlower

2.22. Query

Dalam studi perbandingan ini hanya query untuk retrieve data:

MongoDB:

```
db.imagedescriptions.find({});
```

Prosiding
ANNUAL RESEARCH SEMINAR 2016
6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

MySQL

`select * from imagedescriptions;`

2.23. Tools

Ada beberapa tools yang digunakan dalam studi ini, masing masing diperuntukkan untuk MongoDB dan MySQL.

MongoDB Server

MongoDB Server bisa di download pada URL :

<http://www.mongodb.com/download-center>

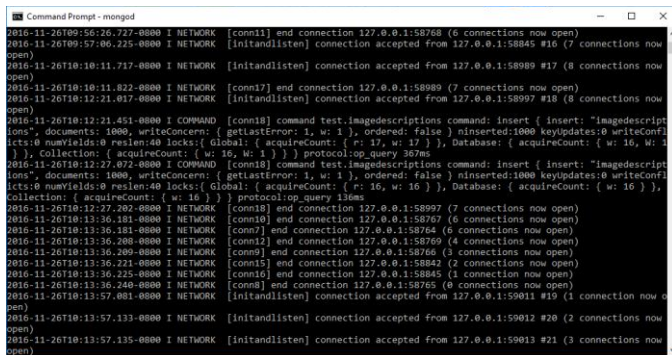


Fig. 10. Mongo Server running dilingkungan Windows

Jika Mongo Server sudah aktif maka aktifitas manajemen database bisa dilaksanakan.

MongoBooster

MongoBooster berfungsi sebagai halaman visual manajemen data untuk MongoDB.

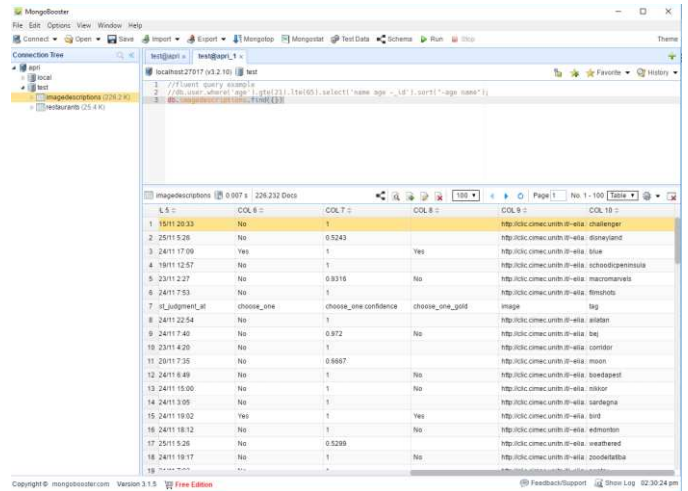


Fig. 11. Halaman kerja MongoBooster

Setelah MongoServer aktif, proses manajemen data lebih mudah dilakukan secara visual, MongoBooster dilengkapi dengan fasilitas export dan import data data sumber data yang berbeda, seperti dari SQL menjadi NoSQL. Bentuk dokumen yang didukung adalah JSON.

XAMPP

XAMPP digunakan untuk menjalankan service MySQL dan PHP pada lingkungan Windows. XAMPP juga berfungsi untuk halaman visual manajemen data.

XAMPP bisa didownload pada URL.

<http://www.apachefriends.org/download.html>

XAMPP menyediakan manajemen data secara visual yang disebut dengan phpMyAdmin.

Figure 3. adalah bentuk visual phpMyAdmin yang sudah diisi dataset penelitian ini.

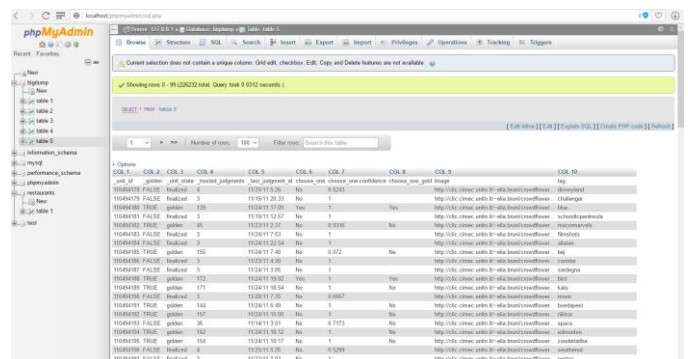


Fig. 12. Halaman kerja phpMyAdmin

2.24. *Lingkungan kerja*

Untuk mendapatkan hasil perbandingan yang adil untuk kedua database maka lingkungan kerja hardware dan software juga memiliki spesifikasi yang sama. Berikut spesifikasi hardware dan sistem operasi yang digunakan.

Hardware:

Processor : Intel(R) Core(TM) i3-4170 CPU @3.70GHz
 RAM : 2.00 GB
 System Type : 64-bit Operating System, x64-based processor

Sistem Operasi:

Windows 10 Pro N

2.25. *Konfigurasi Database*

Sebelum database bisa digunakan baik dilingkungan MongoDB maupun dilingkungan MySQL ada beberapa langkah yang harus dilakukan sebagai berikut:

- Import data ke MongoDB Server

Data yang selesai didownload masih dalam format CSV, harus di import dulu kedalam bentuk JSON [22].

MongoDB memiliki fasilitas import yang bisa langsung digunakan secara command line.

Contoh command line:

```
mongoimport - -db test - -collection imagedescriptions - -type json - -file Image.json - -jsonArray.
```

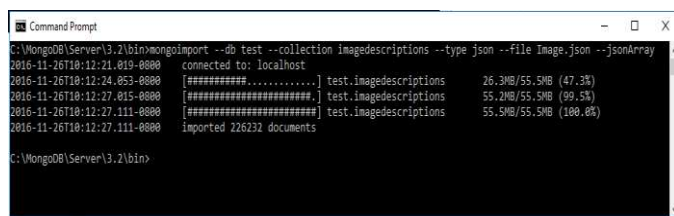


Fig. 13. Proses Impor file JSON ke MongoDB Server.

Setelah proses import data dari format JSON kedalam MongoDB server maka proses manajemen data bisa dilakukan melalui console mongoDB atau juga bisa dilakukan pada MongoBooster, seperti yang terlihat pada Figure 3.

- Import data ke MySQL

Import data ke dalam MySQL bisa melalui phpMyAdmin, proses sangat mudah. Namun karena dalam studi ini menggunakan data berukuran besar, maka memerlukan waktu untuk proses import ke MySQL.

V. HASIL STUDI PERBANDINGAN

Setelah melakukan import data ke masing-masing database, saatnya untuk melakukan studi perbandingan antara kedua basisdata, skenarionya adalah hanya membandingkan satu perintah query saja pada masing masing database. Query yang dilakukan untuk masing-masing database adalah:

MongoDB:

```
db.imagedescriptions.find({});
```

MySQL

```
select * from imagedescriptions;
```

Query diatas berfungsi untuk menampilkan data yang ada dalam database.

Table 1. Merupakan tabel hasil response time masing-masing database setelah diberikan Query.

TABLE VI. HASIL RESPONSE TIME QUERY

MongoDB (dalam satuan detik)	MySQL (dalam satuan detik)
12.83624	70.584384

Table 1. dan Figure 6 memberikan visualisasi tentang hasil yang didapat dalam studi banding performansi antara MongoDB dan MySQL.

Hasil didapat dengan mencatat respon time setelah query dilakukan pada masing masing database.

Dari hasil pencatatan ini dapat disimpulkan bahwa MongoDB memberikan respon time yang cepat dibandingkan MySQL.

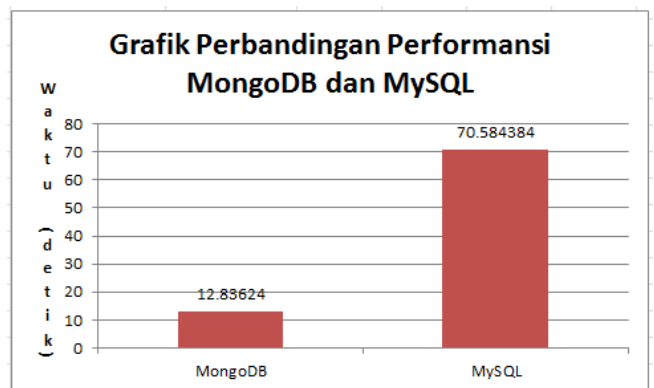


Fig. 14. Proses Impor file JSON ke MongoDB Server.

Prosiding
ANNUAL RESEARCH SEMINAR 2016
6 Desember 2016, Vol 2 No. 1

ISBN : 979-587-626-0 | UNSRI

<http://ars.ilkom.unsri.ac.id>

VI. PENUTUP

2.26. *Kesimpulan*

Dalam studi perbandingan ini didapatkan hasil response time untuk masing masing database, mongoDB menunjukkan performansi yang baik dengan response time 12.83624 detik, MySQL menunjukkan response time 70.584384 detik dengan 226.232 record untuk masing masing database.

Kelebihan studi ini adalah menggunakan dataset yang free dengan jumlah record yang besar.

Kekurangan studi ini adalah hanya menggunakan satu query saja untuk mendapatkan hasil query dan dicatat sebagai hasil perbandingan performansi antara MongoDB dan MySQL.

2.27. *Saran*

Untuk pengembangan penelitian berikutnya diharapkan untuk pembuatan sebuah aplikasi yang mampu menjalankan semua query dan mendapatkan hasil response time.

Aplikasi bisa saja berbentuk tools basis data yang dibangun menggunakan Java, .NET Framework serta aplikasi berbasis web yang dikembangkan dengan PHP atau PHP Framework.

REFERENCES

- [1] R. J. Brunner and E. J. Kim, "Teaching Data Science," *Procedia - Procedia Comput. Sci.*, vol. 80, pp. 1947–1956, 2016.
- [2] I. Abaker, T. Hashem, V. Chang, N. Badrul, K. Adewole, I. Yaqoob, A. Gani, E. Ahmed, and H. Chiroma, "International Journal of Information Management The role of big data in smart city," *Int. J. Inf. Manage.*, vol. 36, no. 5, pp. 748–758, 2016.
- [3] H. Fang, Z. Zhang, C. J. Wang, M. Daneshmand, C. Wang, and H. Wang, "A Survey of Big Data Research," no. October, pp. 6–9, 2015.
- [4] W. Van Der Aalst, S. Member, E. Damiani, and S. Member, "Processes Meet Big Data : Connecting Data Science with Process Science," vol. 8, no. 6, pp. 810–819, 2015.
- [5] M. Bardi, Z. Xianwei, L. I. Shuai, and L. I. N. Fuhong, "Big Data security and privacy : A review," no. 2, pp. 135–145, 2014.
- [6] M. Swan, "Philosophy of Big Data Expanding the Human-Data Relation with Big Data Science Services," 2015.
- [7] J. P. D. Comput, K. Kambatla, G. Kollias, V. Kumar, and A. Grama, "Trends in big data analytics," *J. Parallel Distrib. Comput.*, vol. 74, no. 7, pp. 2561–2573, 2014.
- [8] C. Györödi, R. Györödi, G. Pecherle, and A. Olah, "A Comparative Study : MongoDB vs . MySQL," pp. 0–5, 2015.
- [9] M. A. Mohamed and O. G. Altrafi, "Relational vs . NoSQL Databases : A Survey," vol. 03, no. 03, pp. 598–601, 2014.
- [10] J. Bhogal and I. Choksi, "Handling Big Data using NoSQL," pp. 393–398, 2015.
- [11] M. Katkar, "Performance Analysis for NoSQL and SQL," vol. 2, no. 3, pp. 12–17, 2015.
- [12] H. Zhang, G. Chen, and C. Ooi, "In-Memory Big Data Management and Processing : A Survey," vol. 27, no. 7, pp. 1920–1948, 2015.
- [13] Z. Parker, S. Poe, and S. V. Vrbsky, "Comparing NoSQL MongoDB to an SQL DB," 2013.
- [14] R. Hecht, "NoSQL Evaluation A Use Case Oriented Survey," pp. 336–341, 2011.
- [15] J. Han and J. E. Haihong, Le, Guan, Du, "Survey on NoSQL Database," pp. 363–366, 2011.
- [16] H. Gonz and A. E. Chis, "Cloud-based NoSQL Data Migration," 2016.
- [17] A. M. W. Paper, "Top 5 Considerations When Evaluating NoSQL Databases," no. August, 2015.
- [18] M. Fotache and D. C. Al, "NoSQL and SQL Databases for Mobile Applications. Case Study: MongoDB versus PostgreSQL," vol. 17, no. 2, pp. 41–58, 2013.
- [19] S. Chickerur, "Comparison of Relational Database with Document-Oriented Database (MongoDB) for Big Data Applications," pp. 41–47, 2015.
- [20] R. Cattell, "Scalable SQL and NoSQL Data Stores," vol. 39, no. 4, 2010.
- [21] K. Kaur, R. Rani, C. Sci, and E. Deptt, "Modeling and Querying Data in NoSQL Databases," 2013.
- [22] Y. Liao, J. Zhou, C. Lu, S. Chen, C. Hsu, W. Chen, M. Jiang, and Y. Chung, "Data adapter for querying and transformation between SQL and NoSQL database," *Futur. Gener. Comput. Syst.*