การเปรียบเทียบการทำงานของอัลกอริทึมกับฐานข้อมูลเอสคิวแอลและโนเอสคิวแอล

A Comparison of Algorithms of SQL and NoSQL Databases

สมใจ จิตคำนึงสุข[1], ทิพย์มณฑา ผกาแก้ว[2], พิเชษฎ์ จุลรอด[3]

[1]สาขาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต
21 ถ.เทพกระษัตรี ต.รัษฎา อ.เมือง จ.ภูเก็ต 83000 โทรศัพท์ : 0-76211959
E-mail: somjai.j@pkru.ac.th
[2]สาขาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต
21 ถ.เทพกระษัตรี ต.รัษฎา อ.เมือง จ.ภูเก็ต 83000 โทรศัพท์ : 0-76211959
E-mail: thipmonta.paka@gmail.com
[3]สาขาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏภูเก็ต
21 ถ.เทพกระษัตรี ต.รัษฎา อ.เมือง จ.ภูเก็ต 83000 โทรศัพท์ : 0-76211959
E-mail: phichete@pkru.ac.th

บทคัดย่อ

การนำข้อมูลจำนวนมหาศาล (Big Data) มาใช้ประโยชน์ ประสบปัญหาหลายด้านสำหรับการขยายขีดความสามารถกับระบบการจัดการฐานข้อมูลเชิงสัมพันธ์ (RDBMS) ทำให้การดำเนินการมีประสิทธิภาพลดอย่างรวดเร็วตามปริมาณข้อมูลที่เพิ่มขึ้น แนวโน้มในการกำจัดสิ่งดังกล่าวได้นำไปสู่การพัฒนารูปแบบข้อมูลใหม่ในการใช้ระบบจัดเก็บฐานข้อมูล งานวิจัยนี้ได้นำเสนอฐานข้อมูลที่เป็น NoSQL มาใช้เป็นทางเลือกคือ MongoDB เพื่อลดข้อจำกัดของภาษาสอบถามเชิงโครงสร้าง (SQL) ได้แสดงให้เห็นว่าการใช้งานของ MongoDB เป็นการนำเสนอความสามารถในการเข้าถึงการจัดเก็บ และยังคงสามารถใช้งานร่วมกันกับฐานข้อมูลเชิงสัมพันธ์ จากการวิจัยพบว่า ผลการทดสอบความเร็วในการอ่านข้อมูลของ MongoDB ที่จำนวน 10 คลอสโอเวอร์ (Cross over) เพื่อเข้าถึงข้อมูลจำนวน 48,000 รายการ ใช้เวลาประมวลผลเฉลี่ย 6.42 วินาที ในขณะที่ SQL ใช้เวลาประมวลผลเฉลี่ย 10.25 วินาที สำหรับการแก้ไขและการลบข้อมูลพบว่า MongoDB ประมวลผลเสร็จภายในเวลา 2.50 และ 2.45 วินาที ในขณะที่ SQL ประมวลผลเสร็จภายในเวลา 5.59 และ 5.50 วินาที ตามลำดับ

คำสำคัญ: ข้อมูลจำนวนมาหาศาล, เอสคิวแอล, ฐานข้อมูลแบบไม่มีโครงสร้าง, ฐานข้อมูลมองโก

Abstract

The utilisation of Big Data is confronting scalability issues associated with Relational Database Management System (RDBMS) in which the performance degrades rapidly as data volumes increases. Periodically, several models have been recommended as alternative solutions for addressing the scalability issues. This article proposes a NoSQL database 'MongoDB' for addressing significant limitations of Structured Query Language (SQL). Several key perspectives of MongoDB, including accessibility, storage, and collaboration with relational databases, were also discussed in this article. We found that the data reading speed at 10 crossovers of MongoDB for accessing the data of 48,000 items took an average of 6.42 seconds while the average time

of SQL under the same criteria took 10.25 seconds. Regarding update and deletion speed, MongoDB could process them at 2.50 and 2.45 seconds respectively, which were significantly faster than 5.59 and 5.50 seconds of SQL.

Keywords: Big Data, SQL, NoSQL, MongoDB

1. คำนำ

ปัจจุบันการใช้ข้อมูล และการนำเทคโนโลยีมาใช้จัดการข้อมูล ทั้งด้านธุรกิจ การตลาด การศึกษา สื่อบันเทิง มีข้อมูลขนาดใหญ่ (Big Data) เกิดขึ้นกับการใช้งานข้อมูลที่มีการเปลี่ยนแปลงอย่างรวดเร็วทั้งการเขียนและการอ่านข้อมูล [1] โดยทาง IDC (Internet Data Center) คาดการณ์ไว้ว่า Big Data จะมีขนาดเพิ่มมากขึ้นถึง 40 เซตตาไบท์ภายในปี 2563 ซึ่งเป็นการเติบโตมากกว่า 50 เท่าจากปี 2553 ดังนั้นความสามารถในการดึงข้อมูลที่มีประโยชน์และมีคุณค่าจากฐานข้อมูลขนาดใหญ่ เพื่อใช้ในการประมวลผลและวิเคราะห์ข้อมูลในเชิงธุรกิจ (Business Analytics) จึงเป็นเรื่องที่หลายองค์กรให้ความสำคัญ

จากการจัดอันดับซอฟต์แวร์ใช้งานในการจัดการฐานข้อมูล (DBMS) ที่ได้รับความนิยมที่ผ่านมา (DB-Engines Ranking, 2016) รายงานแนวโน้มการใช้ระบบการจัดการฐานข้อมูลในช่วงระยะ 2 ปี (Mar 2015-Mar 2016) พบว่า MongoDB ได้รับการจัดอันดับให้เป็นระบบจัดการฐานข้อมูลที่รองรับข้อมูลขนาดใหญ่ (Big data) อยู่ในอันดับที่ 4 จาก DBMS ทั้งหมด จำนวน 299 ระบบ [2]

จากแนวโน้มดังกล่าว ผู้วิจัยได้ทำการศึกษารูปแบบการใช้งานและวิจัยทดลองระบบการทำงานการเข้าถึงข้อมูลและการปรับปรุงข้อมูลในรูปแบบต่าง ๆ ของ MongoDB และเปรียบเทียบกับ SQL โดยวัตถุประสงค์เพื่อทดสอบคุณสมบัติเรื่องความเร็วในการประมวลผลของฐานข้อมูล MongoDB ซึ่งเป็นฐานข้อมูลโนเอสคิวแอล โดยเปรียบเทียบความเร็วในการอ่าน (Read) เขียน (Write) แก้ไข (Update) และการลบ (Delete) มีปัจจัยในการทดสอบ คือ จำนวนข้อมูลและจำนวนแอตตริบิวต์ข้อมูลที่แตกต่างกัน โดยงานวิจัยนี้ประกอบด้วยหัวข้อที่ 1 คำนำ หัวข้อที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง หัวข้อที่ 3 วิธีการดำเนินงานวิจัย หัวข้อที่ 4 ผลการวิจัยและอภิปรายผล หัวข้อที่ 5 ข้อเสนอแนะ หัวข้อที่ 6 กิตติกรรมประกาศ และเอกสารอ้างอิง

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ฐานข้อมูลเอสคิวแอล (SQL)

ระบบการจัดการฐานข้อมูลเชิงสัมพันธ์ (RDBMS) จัดเก็บข้อมูลในรูปของตาราง (Table) มีการจัดการข้อมูลโดยใช้ภาษาสอบถามเชิงโครงสร้าง (SQL) [3] ประเภทของคำสั่งในภาษา SQL แบ่งออกเป็น 3 ประเภท คือ 1) ภาษาสำหรับการนิยามข้อมูล (Data Definition Language: DDL) 2) ภาษาสำหรับการจัดการข้อมูล (Data Manipulation Language: DML) 3) ภาษาควบคุม (Data Control Language: DCL)

2.2 โนเอสคิวแอล (NoSQL)

โนเอสคิวแอลภาษาสอบถามแบบไม่มีโครงสร้าง (Not Only SQL: NoSQL) เป็นฐานข้อมูลที่มีลักษณะแบบไม่สัมพันธ์กัน [4] แบ่งลักษณะการจัดเก็บ 4 ประเภท 1) Key-value data stores ฐานข้อมูลที่เก็บข้อมูลในรูปแบบ Key และ Value ที่สัมพันธ์กันอยู่ ซึ่ง Value จะถูกค้นเจอได้โดยค่า Key 2) Column-based databases ฐานข้อมูลที่เก็บข้อมูลในรูปของตารางคล้ายกับฐานข้อมูลเชิงสัมพันธ์แต่จะเก็บข้อมูลในแบบคอลัมน์แทนแบบแถว เหมาะกับการทำเหมืองข้อมูล 3) Document-based databases ฐานข้อมูลที่เก็บข้อมูลในรูปแบบกลุ่มของเอกสารซึ่งง่ายต่อการแก้ไขโครงสร้างโดยแต่ละเอกสารจะมีจำนวนฟิลด์ที่แตกต่างกันได้ 4) Graph-based data-stores คือฐานข้อมูลที่เก็บข้อมูลในรูปแบบของกราฟที่มีความสัมพันธ์กันทำให้สามารถใช้ทฤษฎีกราฟได้ในการจัดการกับข้อมูล

2.3 ข้อมูลจำนวนมหาศาล (Big Data)

Big Data เป็นข้อมูลจำนวนมหาศาลที่เกิดขึ้นในช่วงเวลาที่มีการเปลี่ยนแปลงทางด้านเทคโนโลยี การสื่อสาร มีการใช้อินเทอร์เน็ต สื่อสังคม (Social Media) อย่างแพร่หลาย มีอัตราการเพิ่มขึ้นในลักษณะทวีคูณและรวมเร็ว [5] ประโยชน์ของ Big Data ที่เห็นได้ชัดเจนมีอยู่ 2 ประการ คือ 1) การวิเคราะห์ (Analytical Use) และ 2) การเกิดผลิตภัณฑ์ใหม่ (Enabling New Products) [6-7]

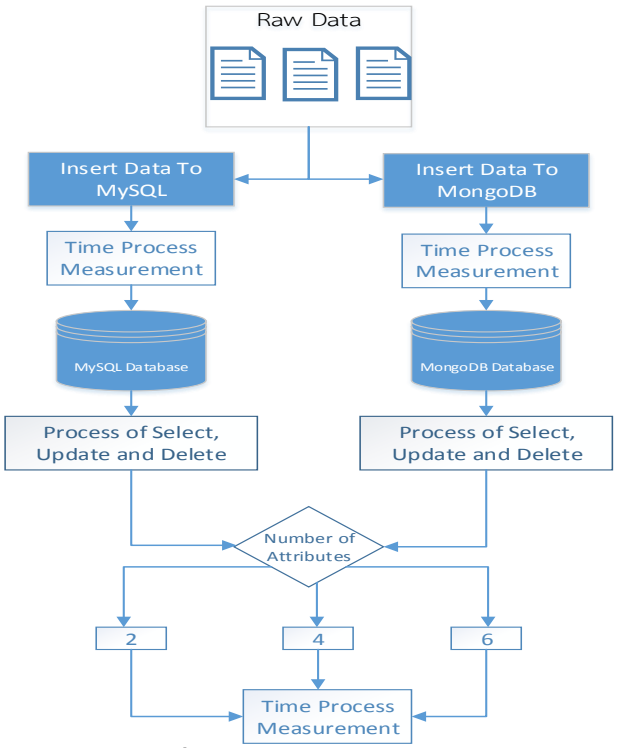
3. วิธีการดำเนินการวิจัย

3.1 การศึกษาและรวบรวมข้อมูล

ผู้วิจัยได้ทำการศึกษา รวบรวมข้อมูล โดยนำข้อมูลของนักศึกษาที่ลงทะเบียนในหมวดของการศึกษาทั่วไป 4 รายวิชา คือ 9901101: ภาษาอังกฤษเพื่อการสื่อสารทั่วไป, 9901102: ภาษาอังกฤษเพื่อการเตรียมตัวสำหรับวิชาชีพ, 9904102: ทักษะการเรียนรู้สารสนเทศ และ 9904104: เทคโนโลยีสารสนเทศเพื่อการเรียนรู้จากฐานข้อมูลมหาวิทยาลัยราชภัฏภูเก็ต [8] ในการลงทะเบียนในรายวิชาดังกล่าว ทั้ง 4 รายวิชา มีจำนวนทั้งสิ้น 48,000 รายการ นำมาวิเคราะห์ตามกระบวนการที่ได้กำหนดไว้

3.2 กรอบแนวคิดของการวิจัย

กรอบแนวคิดสำหรับการวิจัยคือ นำข้อมูลที่ได้มาเพิ่มข้อมูลลงในฐานข้อมูลทั้ง SQL และ MongoDB เพื่อเรียกการทำงาน ด้วยการใช้ PHP Script เป็นตัวประมวลผล และติดต่อกับฐานข้อมูลทั้งสองตัวโดยเขียนคำสั่งในการเรียกข้อมูล รวมทั้ง เพิ่ม ลบ แก้ไขข้อมูล และประมวลผล เก็บรายการทำงานรวมถึงบันทึกเวลาการทำงานแต่ละประเภท สามารถแสดงได้ดังภาพที่ 1



ภาพที่ 1: กรอบแนวคิดในการทำวิจัย

จากภาพกรอบแนวคิดในการทำวิจัย สามารถแบ่งการทำงานออกเป็น 3 ประเด็นหลักของกระบวนการทำงาน คือ 1) ส่วนสำหรับผู้ใช้งานเพิ่มข้อมูลเข้าในระบบด้วย Script PHP ตามข้อมูลดิบที่ได้มาจากฐานข้อมูลของมหาวิทยาลัยราชภัฏภูเก็ต โดยจัดเก็บข้อมูลลงใน MongoDB ที่เป็นฐานข้อมูลไม่สัมพันธ์แบบเอกสาร และใช้สำหรับจัดเก็บข้อมูลการประมวลผลรายการที่ใช้ฐานข้อมูลไม่สัมพันธ์แบบเอกสารเช่นเดียวกับ ในส่วนของ SQL ซึ่งเป็นฐานข้อมูลเชิงสัมพันธ์ เพื่อใช้ในการสืบค้น การเพิ่ม ลบและแก้ไขข้อมูล รวมทั้งการเปลี่ยนตัวเลือกแอทริบิวท์ในการดำเนินการทั้ง 4 กระบวนการ คือ กำหนดขอบเขตในการสืบค้น การเพิ่ม ลบและแก้ไขข้อมูล 2) ส่วนของคำสั่งในการจับเวลาเพื่อเก็บเป็นสถิติในการทำงานของการดำเนินการของข้อมูลตามขอบเขตที่สนใจ และ 3) ส่วนการบันทึกเวลาในการประมวลผลของแต่ละกระบวนการ

3.3 การทดสอบความเร็วในการประมวลผล

การทดสอบความเร็วในการประมวลผล แบ่งการทดสอบโดยใช้ 4 กระบวนการ เริ่มจากการเพิ่ม (Insert) การค้นคืน (Select) การแก้ไข (Update) และการลบ (Delete) ข้อมูลใช้จำนวนแอทริบิวท์และจำนวนข้อมูลที่แตกต่างกัน ซึ่งรายละเอียดการทดสอบความเร็วการประมวลผลสามารถแสดงได้ดัง ตารางที่ 1 โดยการทดสอบความเร็วในการประมวลผลแต่ละวิธีการประมวลผลจะใช้ค่าเฉลี่ยจากการทำงาน 10 รอบ (10 Cross over) เพื่อให้ได้ค่าที่แม่นยำ โดยคำนวณจากเวลาทั้งหมดที่ใช้ในการประมวลผล (Total Time) แต่ละรอบการทำงานซึ่งสามารถแสดงได้ดังสมการที่ 1 หลังจากนั้นนำค่า Total Time(N) มาหาค่าเฉลี่ย (Average_Time) โดยใช้ค่าเฉลี่ยเลขคณิต (Arithmetic Mean) และสมการที่ 2 ตามลำดับ

$$Total_Time(n) = Finished_Time(n) - Start_Time(n) \quad (1)$$

โดยที่ Total_Time(n) คือ เวลาทั้งหมดที่ใช้ในการประมวลผลแต่ละรอบ

Finished_Time(n) คือ เวลาสิ้นสุดที่ใช้ในการประมวลผลแต่ละรอบ

Start_Time(n) คือ เวลาเริ่มต้นที่ใช้ในการประมวลผลแต่ละรอบ

$$Average_Time = Sum(Total_Time(n)) / N : n=1..N \quad (2)$$

ตารางที่ 1: กระบวนการในการทำงานและจำนวนรายการที่ใช้ในการประมวลผล

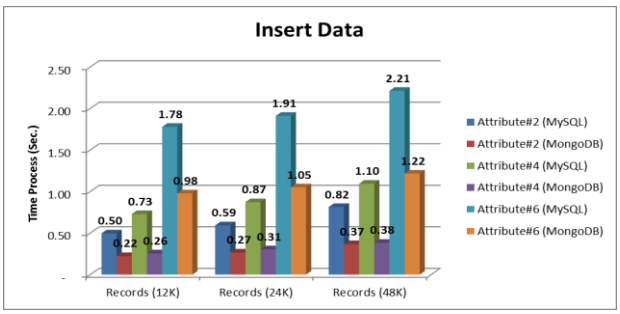
วิธีการประมวลผล	จำนวน 2 แอทริบิวท์			จำนวน 4 แอทริบิวท์			จำนวน 6 แอทริบิวท์		
	จำนวนข้อมูล			จำนวนข้อมูล			จำนวนข้อมูล		
Insert	12,000	24,000	48,000	12,000	24,000	48,000	12,000	24,000	48,000
Select	12,000	24,000	48,000	12,000	24,000	48,000	12,000	24,000	48,000
Update	12,000	24,000	48,000	12,000	24,000	48,000	12,000	24,000	48,000
Delete	12,000	24,000	48,000	12,000	24,000	48,000	12,000	24,000	48,000

4. ผลการวิจัยและอภิปรายผล

ผลการทดสอบความเร็วในการประมวลผลข้อมูลของแต่ละวิธีการดำเนินการ พิจารณาจากจำนวนแอทริบิวต์และเลือกจำแนกข้อมูลเป็นรายการ 3 ระดับ คือ 12,000 รายการ 24,000 รายการ และ 48,000 รายการ โดยทั้ง 3 ระดับการเลือกข้อมูลได้ดำเนินการเหมือนกันในการใช้ฐานข้อมูลจาก SQL และ MongoDB และนำค่าเฉลี่ย 10 ครั้ง นำมาเปรียบเทียบกัน ซึ่งผลการวิจัยมีรายละเอียดดังต่อไปนี้

4.1 ผลการทดสอบความเร็วในการเพิ่มข้อมูล

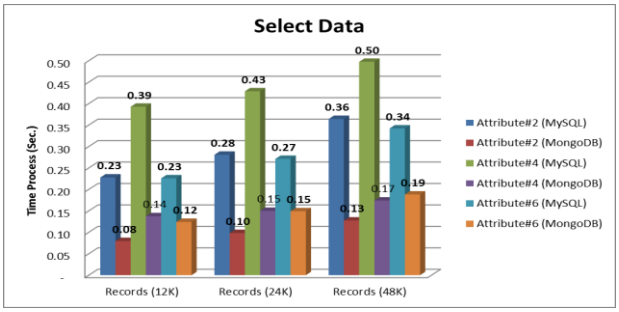
จากผลการทดสอบความเร็วในการเพิ่มข้อมูล (Insert) ของฐานข้อมูล MongoDB และ SQL พบว่าจำนวนรายการข้อมูลที่แตกต่างกัน ฐานข้อมูล MongoDB ใช้เวลาในการประมวลผลน้อยกว่าเฉลี่ยลดลง ร้อยละ 50 โดยประมาณ แต่ละการทดสอบใช้เวลาประมวลผลไม่ถึง 1 วินาที ซึ่งสามารถแสดงได้ดังภาพที่ 2



ภาพที่ 2: การประมวลผลเพื่อเพิ่มข้อมูลของ MongoDB และ SQL

4.2 ผลการทดสอบความเร็วในการอ่านข้อมูล

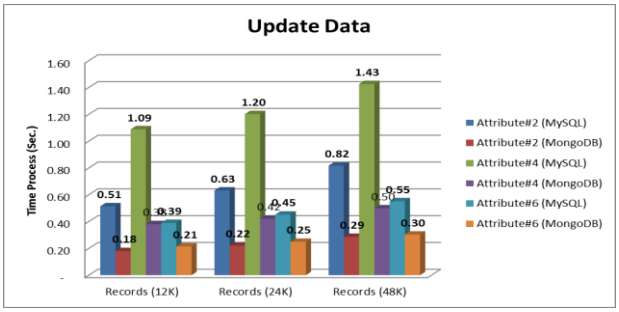
จากผลการทดสอบความเร็วในการอ่านข้อมูล (Select) ของฐานข้อมูล MongoDB พบว่า จำนวนการทำงาน 10 Cross over ที่ระดับข้อมูล 12,000 24,000 และ 48,000 รายการ ใช้เวลาในการอ่านข้อมูลต่างกันมาก อยู่ที่ค่าเฉลี่ย 0.08 ถึง 0.19 วินาที ในขณะที่ SQL มีค่าเฉลี่ยอยู่ที่ 0.23 ถึง 0.50 ซึ่งสามารถแสดงได้ดังภาพที่ 3



ภาพที่ 3: การประมวลผลเพื่ออ่านข้อมูลของ MongoDB และ SQL

4.3 ผลการทดสอบความเร็วในการแก้ไขข้อมูล

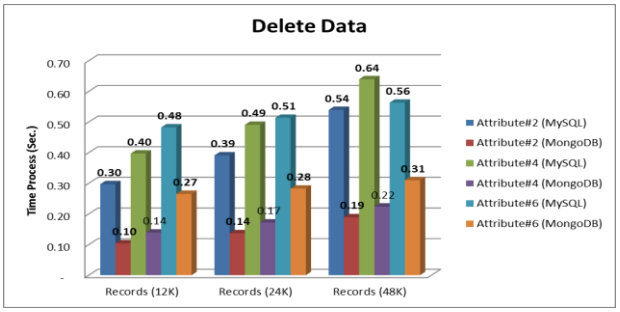
จากผลการทดสอบความเร็วในการแก้ไขข้อมูล (Update) ของฐานข้อมูล MongoDB พบว่า จำนวนการทำงาน 10 Cross over ที่ระดับข้อมูล 12,000 24,000 และ 48,000 รายการ ใช้เวลาในการแก้ไขข้อมูลต่างกันมาก อยู่ที่ค่าเฉลี่ย 0.18 ถึง 0.50 วินาที ในขณะที่ SQL มีค่าเฉลี่ยอยู่ที่ 0.39 ถึง 1.43 ซึ่งสามารถแสดงได้ดังภาพที่ 4



ภาพที่ 4: การประมวลผลเพื่อแก้ไขข้อมูลของ MongoDB และ SQL

4.4 ผลการทดสอบความเร็วในการลบข้อมูล

จากผลการทดสอบความเร็วในการลบข้อมูล (Delete) ของฐานข้อมูล MongoDB และ SQL พบว่าจำนวนรายการข้อมูลที่แตกต่างกัน ฐานข้อมูล MongoDB ใช้เวลาในการประมวลผลน้อยกว่าเฉลี่ยลดลง ร้อยละ 45 โดยประมาณ แต่ละการทดสอบใช้เวลาประมวลผลไม่ถึง 0.5 วินาที ซึ่งสามารถแสดงได้ดังภาพที่ 5



ภาพที่ 5: การประมวลผลเพื่อลบข้อมูลของ MongoDB และ SQL

4.5 อภิปรายผลการทดลอง

จากผลการทดสอบความเร็วในเพิ่มข้อมูล การอ่านข้อมูล การแก้ไข และการลบข้อมูล ของฐานข้อมูล MongoDB และ SQL พบว่าการเพิ่มข้อมูล MongoDB มีค่าเฉลี่ยน้อยกว่าการใช้ SQL เฉลี่ย 25% การเข้าถึงข้อมูล MongoDB มีค่าเฉลี่ยน้อยกว่า การใช้ SQL เฉลี่ย 35% การแก้ไขข้อมูล MongoDB มีค่าเฉลี่ยน้อยกว่า การใช้ SQL เฉลี่ย 37% การลบข้อมูล MongoDB มีค่าเฉลี่ยน้อยกว่า การใช้ SQL เฉลี่ย 50% ตามลำดับ จากผลดังกล่าวจะเห็นได้ว่าการใช้ MongoDB ซึ่งเป็น NoSQL เป็นอีกทางเลือกสำหรับการจัดเก็บข้อมูลที่ใช้สำหรับการใช้ข้อมูล Big Data และจะเป็นประโยชน์อย่างยิ่งในอนาคตที่จะทำเหมืองข้อมูลและวิจัยข้อมูลในเชิงลึกต่อไป

5. ข้อเสนอแนะ

จากการทำวิจัยในครั้งนี้มีข้อเสนอแนะเพื่อระบบทำงานของการใช้ฐานข้อมูล MongoDB ควรใช้ระบบประมวลผลแบบ 64 บิต ซึ่งสามารถใช้งานร่วมกับไดรเวอร์ของ MongoDB ได้สะดวกกว่า และงานวิจัยในอนาคตจะทำการทดสอบเปรียบเทียบ NoSQL มากกว่า 1 ชนิด รวมถึงทดสอบความแม่นยำในการค้นคืนข้อมูลนอกเหนือจากการทดสอบความเร็วในการประมวลผลเพียงอย่างเดียว

6. กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณ สำนักวิทยบริการและเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏภูเก็ตที่สนับสนุนและให้ความอนุเคราะห์ข้อมูลเพื่อใช้ในการทำวิจัยในครั้งนี้

เอกสารอ้างอิง

[1] T. Tunguz, “The Hottest Startup Sectors In 2016”, [Online]. Available: http://seekingalpha.com/article/3789176-hottest-startup-sectors-2016. [Accessed: 17-Mar-2016].

[2] DB-Engines, “DB-Engines Ranking - Trend Popularity”, 2016. [Online]. Available: http://db-engines.com/en/ranking_trend. [Accessed: 4-Mar-2016].

[3] สุจิตรา อดุลย์เกษม, *ระบบฐานข้อมูล (Database System)*, (พิมพ์ครั้งที่ 1). กรุงเทพฯ: ท้อป, 2553.

[4] ผุสดี บุญรอด และ ประกายมาศ ศรีสุขทักษิณ, “การค้นคืนข้อมูลขนาดใหญ่โดยใช้ภาษาสอบถามแบบไม่มีโครงสร้างร่วมกับเทคโนโลยีเว็บเชิงความหมาย” *วารสารวิชาการพระจอมเกล้าพระนครเหนือ* ปีที่ 25 ฉบับที่ 2 พ.ค.-ส.ค. 2558 หน้า 255-264.

[5] ศรีสมรัก อินทุจันทร์ยง, “Big Data จำเป็นสำหรับทุกองค์การหรือไม่ Whether Big Data is Essential for all Organizations or Not,” *วารสารบริหารธุรกิจ*, ปีที่ 38(146), 2558 หน้า.1-8.

[6] E. Dumbill, “What is big data? An introduction to the big data landscape”, 2012. [Online]. Available: http://radar.oreilly.com/2012/01/What-is-big-data.html. [Accessed: 24-Jan-2016].

[7] Panasas, “BigData”, 2012. [Online]. Available: http://www.panasas.com/solutions/big-data. [Accessed: 24-Jan-2016].

[8] สำนักวิทยบริการและเทคโนโลยีสารสนเทศ, “ฐานข้อมูลงานทะเบียนนักศึกษา มหาวิทยาลัยราชภัฏภูเก็ต”, 2559.