

53314314: สาขาวิชาวิทยาการคอมพิวเตอร์และสารสนเทศ

คำสำคัญ: ข้อมูลการท่องเที่ยว / Java Concurrency / Map Reduce

ยุวธิดา ชิวปรีชา: การพัฒนาระบบค้นหาด้วยฮอนโทโลยีบน Hadoop Cluster :

กรณีศึกษาข้อมูลการท่องเที่ยว. อาจารย์ที่ปรึกษาวิทยานิพนธ์: รศ.ดร.จันทนา จันทราพรชัย. 84 หน้า.

การค้นหาข้อมูลการท่องเที่ยวเป็นสิ่งสำคัญสำหรับนักท่องเที่ยว ซึ่งปัจจุบันมีผู้ทำการรวบรวมฮอนโทโลยีเกี่ยวกับการท่องเที่ยวเพื่อนำมาช่วยในการพัฒนาระบบสืบค้นให้มีประสิทธิภาพมากยิ่งขึ้น โดยในงานวิจัยฉบับนี้ได้พัฒนาระบบเพื่อช่วยในการสืบค้นข้อมูลการท่องเที่ยวโดยการนำฮอนโทโลยีที่มีอยู่มาใช้ในการสืบค้นและพยายามเพิ่มประสิทธิภาพการสืบค้นข้อมูลขนาดใหญ่ โดยการทำงานแบบขนาน มีการประมวลผลแบบกระจาย โดยใช้เทคนิคของ MapReduce ใน Hadoop มาช่วยกระจายงานไปประมวลผลเพื่อลดระยะเวลาในการสืบค้นและเพิ่มประสิทธิภาพในการสืบค้นข้อมูลการท่องเที่ยว

จากการทดลองพบว่าในการสืบค้นข้อมูลบนคอมพิวเตอร์เครื่องเดียวนั้น หากใช้รูปแบบการกระจายข้อมูลที่มีการนำข้อมูลเข้าสู่หน่วยความจำให้มากที่สุดแล้วค้นหา จะทำให้ใช้เวลาในการค้นหาเร็วยิ่งขึ้น และในกรณีที่สืบค้นข้อมูลโดยให้คอมพิวเตอร์หลายเครื่องช่วยกันประมวลผลนั้น ยิ่งมีการเพิ่มจำนวนเครื่องและเพิ่มแกนในการประมวลผลจะใช้เวลาในการค้นหาน้อยลง นอกจากนี้ขนาดของการแบ่ง Block ข้อมูลมีผลต่อเวลาในการทำงาน โดยมีผลกับจำนวน Map Task ที่เกิดขึ้น หากมีการเพิ่มขนาด Block เป็น 512 M และกำหนด Reduce Task = 2 บนคอมพิวเตอร์หลายเครื่อง และแต่ละเครื่องมีหลายแกน ประสิทธิภาพการค้นหาจะดีที่สุด แต่หากมีจำนวน Block Size เกิน 512 M ประสิทธิภาพการค้นหาจะเริ่มลดลง ทั้งนี้ขึ้นอยู่กับประสิทธิภาพของเครื่องที่ใช้ในการทดสอบด้วย

---

ภาควิชาคอมพิวเตอร์

บัณฑิตวิทยาลัย มหาวิทยาลัยศิลปากร

ลายมือชื่อนักศึกษา.....

ปีการศึกษา 2557

ลายมือชื่ออาจารย์ที่ปรึกษาวิทยานิพนธ์ .....

53314314: MAJOR: COMPUTER AND INFORMATION SCIENCE

KEY WORD: TOURISM INFORMATION / JAVA CONCURRENCY / MAP REDUCE

YUWATHIDA CHIWPREECHAR: DEVELOPMENT OF SEARCH ENGINE USING ONTOLOGY ON HADOOP CLUSTER : CASE STUDY TOURISM DOMAIN. THESIS ADVISOR: ASSOC.PROF.CHANTANA CHANTRAPORNCHAI,Ph.D. 84 pp.

Tourism data searching is important for tourists. There are efforts to collect tourism ontology to help develop a more powerful search system. This research aims to develop a system to help in the tourism data search by bringing existing ontology technology used in queries and enhancing with parallel processing. Distributed computing using software framework, Hadoop MapReduce is used to process the search task to increase efficiency of tourism search system respectively.

The experimental results show that to search data on a single computer, if all data is in memory, and are distributed properly. The search time can be improved. In the case of search data on multiple multi-core computers can even speedup the search time. Also, how to divide data is also an important factor to the search time. The block size is used to determine the number of map tasks. If the data is divided using the block size equal to 512 Mbytes and using the reduce task = 2 on multiple multi-core computers, it will yield the best search performance, upto 10 times, for searching totally 40G data. If the number of block size exceeds 512Mbytes, it may slow down the search. This, however, depends on the performance of the computer used in the experiment.

---

Department of Computing

Graduate School, Silpakorn University

Student's signature .....

Academic Year 2014

Thesis Advisor's signature .....