

ข้อเสนอโครงการวิจัย
สำนักคอมพิวเตอร์และเทคโนโลยีสารสนเทศ มจพ.
ประจำปีงบประมาณ 2556

1. ชื่อโครงการ/ผลงาน การออกแบบและพัฒนาระบบวิเคราะห์ข้อมูลจราจรเว็บเซิร์ฟเวอร์โดยใช้ HBASE บนพื้นฐานการเก็บข้อมูลแบบ HDFS
(อังกฤษ)
2. ชื่อ-นามสกุล หัวหน้าโครงการนายจันทพงษ์...บุตรลักษณ์.....
ตำแหน่งนักวิชาการคอมพิวเตอร์..... หน่วยงานต้นสังกัดสำนักคอมพิวเตอร์และเทคโนโลยีสารสนเทศ.....
โทรศัพท์ (ที่สามารถติดต่อได้) มือถือ086-1519705..... อีเมล..jantapongb@kmutnb.ac.th.....
3. ประเภทบุคลากร ☐ ข้าราชการ ☒ พนักงานมหาวิทยาลัย
4. ระยะเวลาในการปฏิบัติงาน7..... ปี เดือน
5. ปัจจุบันกำลังศึกษาต่อในระดับ ☐ปริญญาตรี ☐ปริญญาโท ☐ปริญญาเอก ☒ไม่ได้กำลังศึกษาต่อในทุกระดับ
6. ประเภทงานวิจัย ☐ฮาร์ดแวร์หรือสิ่งประดิษฐ์ ☒โปรแกรมซอฟต์แวร์ ☒งานบริหารจัดการ ☐งานอื่นๆ
7. ลักษณะโครงการวิจัย (ตอบได้มากกว่า 1 ข้อ)
☒ สอดคล้องกับยุทธศาสตร์การวิจัยของสำนักคอมพิวเตอร์และเทคโนโลยีสารสนเทศหรือมหาวิทยาลัย
☒ การพัฒนาเทคโนโลยีสารสนเทศ ☒ การสร้างองค์ความรู้ใหม่ หรือขยายฐานความรู้เดิม
☐ การพัฒนาการเรียนการสอน ☐ การพัฒนาด้านบริหารธุรกิจ สังคม เศรษฐกิจ และอุตสาหกรรม
☒ การพัฒนาด้านการบริหารจัดการ / พัฒนาระบบการทำงานสำนักคอมพิวเตอร์และเทคโนโลยีสารสนเทศ
☐ การประดิษฐ์คิดค้นชิ้นงานวิชาการทางด้านเทคโนโลยีสารสนเทศ
8. คณะผู้วิจัย

ชื่อ – สกุล	หน่วยงานต้นสังกัด / โทรศัพท์	ระดับการศึกษา	สัดส่วน	ลายมือชื่อ
.....				
.....				
.....				
9. การติดค้างการส่งรายงานวิจัยฉบับสมบูรณ์ที่ได้รับทุนอุดหนุนทั้งภายในและภายนอกมหาวิทยาลัย ซึ่งสิ้นสุดระยะเวลาตามที่สัญญากำหนดไว้แล้ว ☐ติดค้าง ☒ไม่ติดค้าง
10. ระยะเวลาใช้ในการวิจัย12..... เดือน
11. งบประมาณตลอดโครงการ100,000..... บาท

ลงชื่อ หัวหน้าโครงการ/ผู้ขอรับทุน

(นายจันทพงษ์ บุตรลักษณ์)

ลงชื่อ หัวหน้าฝ่าย/หัวหน้าสำนักงาน

(นายอิสเรศ สมณะ)

ลงชื่อ รองผู้อำนวยการ/ผู้บังคับบัญชาชั้นต้น

(ผศ.ดร.ชูพันธุ์ รัตนโกศา)

แบบเสนอโครงการวิจัย

เพื่อขอรับทุนสนับสนุนการวิจัย สำนักคอมพิวเตอร์และเทคโนโลยีสารสนเทศ

1. **ชื่อผลงานวิจัย** การออกแบบและพัฒนาระบบวิเคราะห์ข้อมูลจราจรเว็บเซิร์ฟเวอร์โดยใช้ HBASE บนพื้นฐานการเก็บข้อมูลแบบ HDFS
2. **ชื่อหัวหน้าโครงการ** นายจันทพงษ์ บุตรลักษณ์
คุณวุฒิ/สาขา วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ
ตำแหน่ง นักวิชาการคอมพิวเตอร์
ความชำนาญ/ประสบการณ์ ระบบแจ้งเตือนการบุกรุก(IDS), ระบบเว็บเซิร์ฟเวอร์ และระบบเซิร์ฟเวอร์เสมือน (Virtualization Server)

3. **หลักการและเหตุผล**

เนื่องจากพระราชบัญญัติว่าด้วยการกระทำความผิดเกี่ยวกับคอมพิวเตอร์ พ.ศ.2550 มาตรา 26 ผู้ให้บริการเครือข่ายคอมพิวเตอร์ต้องเก็บรักษาข้อมูลจราจรทางคอมพิวเตอร์ไว้ไม่น้อยกว่าเก้าสิบวัน ซึ่งผู้ให้บริการจะต้องเก็บรักษาข้อมูลของผู้ใช้บริการเท่าที่จำเป็นเพื่อให้ระบุตัวผู้ให้บริการได้ ดังนั้นภาระจึงเกิดกับผู้ให้บริการเครือข่ายคอมพิวเตอร์ เนื่องด้วยการเก็บข้อมูลที่มีขนาดใหญ่จำเป็นต้องใช้อุปกรณ์ประเภท Network-attached storage (NAS) หรือ Storage area Network (SAN) ที่มีราคาสูง และกรณีที่ต้องการค้นหาข้อมูลของผู้ที่กระทำความผิดต้องใช้เวลาอันยาวนานในการสืบค้น เนื่องจากข้อมูลที่เก็บไว้มีขนาดใหญ่

ปัญหาเกี่ยวกับการจัดการข้อมูลขนาดใหญ่ กำลังได้รับความสนใจเป็นอย่างมาก เนื่องจากข้อมูลต่างๆ ที่จัดเก็บมีความสำคัญเพื่อใช้ในการสืบค้น วิเคราะห์ รวมทั้งประมวลผลข้อมูล อีกทั้งข้อมูลที่จัดเก็บยังมีขนาดใหญ่ขึ้นเรื่อยๆ ทำให้การจัดการกับข้อมูลเหล่านั้นมีความลำบากและยุ่งยากมากขึ้น หากจะเลือกข้อมูลเพื่อจัดเก็บ สามารถเลือกได้จากความรู้หรือปัจจัยที่สำคัญ ณ เวลานั้นๆ แต่พอเวลาผ่านไปปัจจัยที่สำคัญก็เปลี่ยนไปตาม จึงทำให้ค่อนข้างยากที่จะคาดการณ์อนาคต ซึ่งอาจจะก่อให้เกิดปัญหาตามมา เช่น ข้อมูลที่เก็บไว้ใช้ไม่ได้หรือไม่เพียงพอเพื่อจะใช้วิเคราะห์ ดังนั้นจำเป็นที่จะต้องจัดเก็บทุกอย่างไว้ ปัจจุบันได้มีการนำเทคโนโลยีที่เรียกว่า HDFS (Hadoop Distributed File System) มาประยุกต์ใช้งานในการเก็บข้อมูลขนาดใหญ่ หลักการทำงานของ HDFS คือแบ่งข้อมูลขนาดใหญ่ที่ต้องการจะเก็บออกเป็นส่วนย่อยๆ แล้วกระจายข้อมูลส่วนย่อยๆ นั้นไปยังเครื่องคอมพิวเตอร์ต่างๆ ที่เชื่อมต่อ ส่วนของการค้นหาข้อมูลภายในเทคโนโลยี HDFS วิธีที่เหมาะสมที่สุดเรียกว่าวิธี Map/Reduce ซึ่งเป็นการส่งคำสั่งค้นหากระจายไปยังเครื่องคอมพิวเตอร์ทุกเครื่องในระบบ โดยที่ไม่จำเป็นต้องมีการย้ายข้อมูลระหว่างการประมวลผล จึงทำให้สามารถค้นหาข้อมูลได้อย่างรวดเร็ว อย่างไรก็ตาม HDFS เหมาะที่จะเก็บไฟล์แบบ Batch เท่านั้น ซึ่งการบริหารจัดการค่อนข้างลำบาก และการอ่านเขียนไม่เป็น Real-Time อีกด้วย

จากปัญหาและข้อจำกัดบางอย่างของ HDFS ดังกล่าวข้างต้น งานวิจัยนี้ต้องการนำเสนอการออกแบบและพัฒนาระบบวิเคราะห์ข้อมูลจราจรเว็บเซิร์ฟเวอร์โดยใช้ HBASE บนพื้นฐานการเก็บข้อมูลแบบ HDFS โดยพัฒนาวิธีการจัดเก็บข้อมูลจราจรเว็บเซิร์ฟเวอร์บน HDFS และความสามารถในการเข้าถึงข้อมูลแบบ Random, Read/Write และ Real-Time ได้ ซึ่งจะทำให้การบริหารและจัดการฐานข้อมูลขนาดใหญ่เป็นไปอย่างรวดเร็วและมีประสิทธิภาพมากยิ่งขึ้น

4. **วัตถุประสงค์**

- 4.1 เพื่อศึกษาวิธีการจัดเก็บข้อมูลขนาดใหญ่ด้วย HDFS
- 4.2 เพื่อศึกษาวิธีการใช้งานข้อมูลที่จัดเก็บด้วย HBASE
- 4.3 เพื่อพัฒนาระบบจัดการข้อมูลจราจรเว็บเซิร์ฟเวอร์บนฐานข้อมูลขนาดใหญ่วิเคราะห์ประสิทธิภาพจากงานของการเก็บข้อมูลแบบกระจาย

5. **ขอบเขตของการวิจัย**

- 5.1 ระบบที่พัฒนาสามารถจัดการกับข้อมูลจราจรเว็บเซิร์ฟเวอร์ Apache เท่านั้น

- 5.2 ระบบสามารถวิเคราะห์ข้อมูลได้ดังนี้ จำนวนการเข้าใช้เว็บไซต์ในแต่ละวัน แต่ละเดือน และแต่ละปี
- 5.3 ระบบสามารถแสดงผลเป็นกราฟได้ตามการวิเคราะห์ข้อมูล
- 5.4 ระบบสำหรับวิจัยมี 1 NameNode และ 10 DataNode (คอมพิวเตอร์ทั้งหมด 11 เครื่อง)
- 5.5 ข้อมูลที่ใช้สำหรับทดสอบ 1-100 ล้านเรคคอร์ด

6. ประโยชน์ที่คาดว่าจะได้รับ

- 6.1 ได้วิธีการในการจัดการกับข้อมูลขนาดใหญ่ เพื่อเป็นแนวทางในการพัฒนาระบบใช้งานด้านอื่นๆ ต่อไป
- 6.2 ได้ระบบวิเคราะห์ข้อมูลสำหรับเว็บเซิร์ฟเวอร์ในด้านสถิติการใช้งานเว็บไซต์ เพื่อเป็นข้อมูลสำหรับปรับปรุงระบบการให้บริการได้ดียิ่งขึ้น
- 6.3 ได้ความรู้และความเข้าใจสำหรับการจัดการกับข้อมูลขนาดใหญ่ และเป็นข้อมูลให้คนอื่นๆ ได้ศึกษาพัฒนาต่อไป

7. การสืบค้นวรรณกรรม

--

8. วิธีดำเนินการวิจัยพร้อมแผนงาน

ตาราง แผนงาน/กิจกรรม และผลงาน

แผนงาน/กิจกรรม	เดือน											
	มิ.ย	ก.ค	ส.ค	ก.ย	ต.ค	พ.ย	ธ.ค	ม.ค	ก.พ	มี.ค	เม.ย	พ.ค
1. ค้นคว้าและศึกษาข้อมูล	←											
2. ติดตั้งและพัฒนา ระบบ				←					→			
3. ทดสอบและสรุป ผลงานวิจัย								←		→		
4. จัดทำเอกสารสรุป								←				→

9. ระยะเวลาที่ใช้ในการวิจัย

มิถุนายน 2556 ถึง พฤษภาคม 2557

10. งบประมาณในการดำเนินงานวิจัย

10.1 หมวดค่าจ้าง

- ค่าจ้างค้นหาข้อมูลและแปลเอกสารเพิ่มเติม 10,000 บาท
- ค่าจ้างวิเคราะห์ข้อมูลทางสถิติและสรุปผล 5,000 บาท
- ค่าจ้างพิมพ์และบันทึกข้อมูล 2,000 บาท
- ค่าจ้างเขียนไลบรารี (API, Class, Framework) สำหรับพัฒนาระบบ 35,000 บาท

10.2 หมวดค่าตอบแทน

- นายจันทพงษ์ บุตรลักษณ์ 40,000 บาท

10.3 หมวดค่าใช้สอย

- ค่าถ่ายเอกสาร 1,000 บาท

10.4 หมวดค่าวัสดุ

- ค่าจัดทำรูปเล่มวิจัยฉบับสมบูรณ์ 2,000 บาท
 - ค่าเอกสารและหนังสือ 1,500 บาท
- 10.6 *หมวดค่าใช้จ่ายอื่นๆ*
- อื่นๆ เช่น ค่าเดินทางไปค้นคว้าข้อมูลต่างมหาวิทยาลัย, หอสมุดแห่งชาติ ฯลฯ 3,500 บาท

11. เอกสารอ้างอิง

ผศ.ดร. ชูพันธุ์ รัตนโกศา. "การออกแบบและพัฒนาระบบค้นหาข้อมูลจราจรทางคอมพิวเตอร์ด้วยวิธี Map/Reduce บนกรอบการทำงานของ Hadoop" วารสารวิชาการเทคโนโลยีอุตสาหกรรม. ฉบับที่ 3 : กันยายน - ธันวาคม 2555.

Apache Software Foundation. Apache Hadoop Start. [ออนไลน์]. เข้าถึงได้จาก : <http://hadoop.apache.org/>. (วันที่ค้นข้อมูล : 1 กุมภาพันธ์ 2556)

Cloudera.com. Cloudera Enterprise Free. [ออนไลน์]. เข้าถึงได้จาก : <http://www.cloudera.com/content/cloudera/en/products/cloudera-enterprise-free.html/>. (วันที่ค้นข้อมูล : 1 กุมภาพันธ์ 2556)

12. ประวัติผู้วิจัย

ชื่อ นายจันทพงษ์ บุตรลักษณ์

ตำแหน่ง นักวิชาการคอมพิวเตอร์

วุฒิการศึกษา วิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ

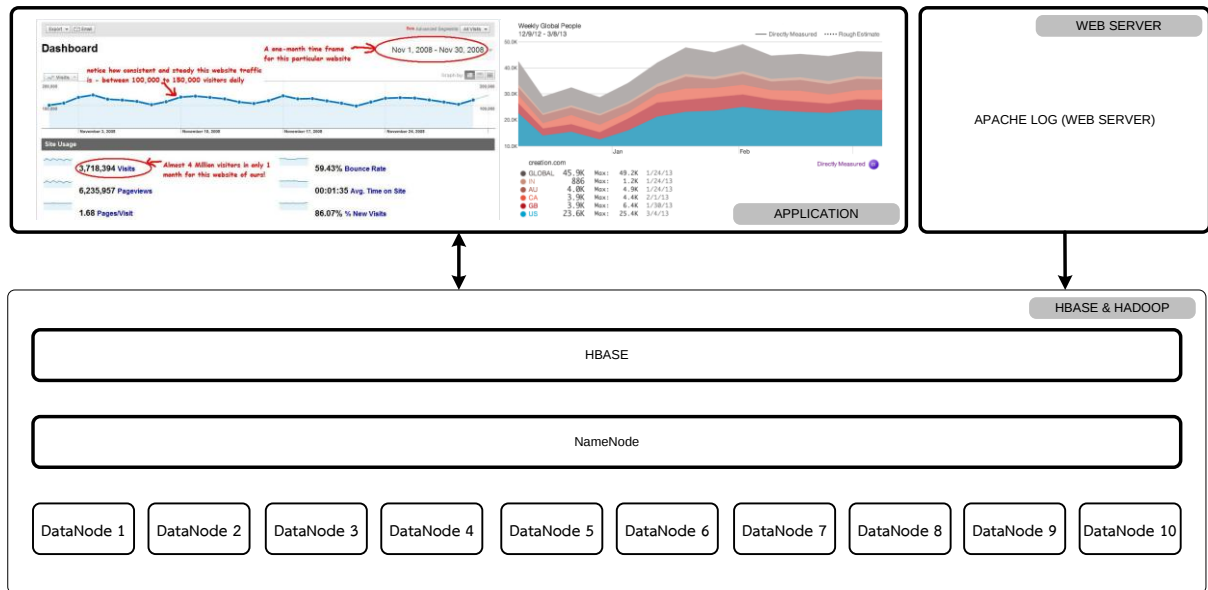
ความชำนาญ/ประสบการณ์ ระบบแจ้งเตือนการบุกรุก, ระบบเว็บเซิร์ฟเวอร์, ระบบเซิร์ฟเวอร์เสมือน, Cloud Computing.

ลงชื่อ หัวหน้าโครงการ

(นายจันทพงษ์ บุตรลักษณ์)

วันที่ เดือน พ.ศ.

ภาพรวมของระบบ



ทรัพยากรสำหรับการทำวิจัย

1. ขอยืมเครื่องคอมพิวเตอร์จำนวน 11 เครื่องจากฝ่ายบริการเพื่อใช้ในการทำการวิจัย
2. ขอใช้พื้นที่จากฝ่ายบริการเพื่อใช้สำหรับจัดวางเครื่องคอมพิวเตอร์ 11 เครื่อง