โปรแกรมตรวจพินิจการใช้ทรัพยากรระบบสารสนเทศอัตโนมัติ

AUTOMATE RESOURCE CONSUMPTION INSPECTION PROGRAM

ณัฐชัย พิมพ์สวัสดิ์
นลินพร ศานติคณาวงศ์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง
ปีการศึกษา 2555

โปรแกรมตรวจพินิจการใช้ทรัพยากรระบบสารสนเทศอัตโนมัติ

AUTOMATE RESOURCE CONSUMPTION INSPECTION PROGRAM

ณัฐชัย พิมพ์สวัสดิ์

นลินพร ศานติคณาวงศ์

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2555

ปริญญานิพนธ์ปีการศึกษา 2555

สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง โปรแกรมตรวจพินิจการใช้ทรัพยากรระบบสารสนเทศอัตโนมัติ

AUTOMATE RESOURCE CONSUMPTION INSPECTION PROGRAM

ผู้จัดทำ

1. นายณัฐชัย	พิมพ์สวัสดิ์	รหัสนักศึกษา	52010319
2. นางสาวนลินพร	ศานติคณาวงศ์	รหัสนักศึกษา	52010581

..... อาจารย์ที่ปรึกษา

(ดร. ธนัญชัย ตรีภาค)

โปรแกรมตรวจพินิจการใช้ทรัพยากรระบบสารสนเทศอัตโนมัติ

นาย ณัฐชัย	พิมพ์สวัสดิ์	52010319
นางสาว นลินพร	ศานติคณาวงศ์	52010581
ดร. ธนัญชัย	ตรีภาค	อาจารย์ที่ปรึกษา

ปีการศึกษา 2555

บทคัดย่อ

ในปัจจุบันมีการนำระบบสารสนเทศมาใช้งานอย่างแพร่หลาย ความต้องการระบบสารสนเทศที่มีการทำงานซับซ้อนมากขึ้นส่งผลให้ระบบสารสนเทศจำเป็นต้องมีขนาดใหญ่ และมีองค์ประกอบที่หลากหลายมากขึ้น จึงทำให้การดูแลระบบสารเทศทำได้ยากขึ้น โดยเฉพาะการตรวจสอบว่าการทำงานใดที่มีการใช้งานทรัพยากรระบบสูงมากจนผิดปกติ และส่งผลต่อระบบงานโดยรวม

โครงงานนี้มีวัตถุประสงค์เพื่อสร้างเป็นโปรแกรมสำหรับค้นหาเหตุการณ์ที่เป็นสาเหตุให้มีการใช้งานทรัพยากรในระบบสารสนเทศสูงมากจนผิดปกติ โดยการรวบรวมข้อมูลการทำงานของระบบสารสนเทศทั้งหมด และข้อมูลการใช้งานทรัพยากรของอุปกรณ์ต่าง ๆ ภายในระบบมาเก็บไว้ที่แหล่งเดียวกัน แล้วนำมาทำการวิเคราะห์โดยใช้เทคนิคทางสถิติ เพื่อหาความสัมพันธ์ระหว่างข้อมูลการทำงานกับข้อมูลการใช้งานทรัพยากร แล้วนำมาวิเคราะห์ ประมวลผลเพื่อค้นหารายละเอียดการทำงานที่มีการใช้งานทรัพยากรมากผิดปกติ พร้อมทั้งทำการแสดงผลลัพธ์เพื่อให้ผู้ดูแลระบบสามารถนำไปใช้ตรวจสอบระบบสารสนเทศได้โดยง่าย

AUTOMATE RESOURCE CONSUMPTION INSPECTION PROGRAM

Mr. Nattachai	Pimsawad	52010319
Ms. Nalinporn	Santikanawong	52010581
Dr. Thanunchai	Threepak	Advisor

Academic Year 2012

ABSTRACT

Information systems are widely used to support recent business operation. Since many organizations have to achieve complicated functions, current information systems always have a large size and a variety of elements. System administrators have to work more difficult. Especially in finding the application requests that cause system problem

The purpose of this project is to create automate resource consumption inspection program. To achieve all tasks, system log and resource utilization data are collected and analyzed by using business intelligence techniques. Then, results from analysis method are brought to summarize in web interface for easy use.

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้เนื่องจากได้รับคำแนะนำ และคำปรึกษาในการทำวิจัย ศึกษา ค้นคว้า จากดร.ธนัญชัย ตรีภาค อาจารย์ที่ปรึกษาในการปริญญานิพนธ์ จึงทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี กลุ่มผู้วิจัยจึงขอขอบพระคุณอาจารย์เป็นอย่างสูง รวมไปถึงห้องวิจัยและพัฒนาการรักษาความปลอดภัยข้อมูล (ISAG) สาขาวิศวกรรมศาสตร์คอมพิวเตอร์และอาจารย์ที่ปรึกษาประจำห้องวิจัยที่เอื้อเฟื้อสถานที่ทำงาน และสนับสนุนในส่วนของเครื่องมือประกอบการทำวิจัย อีกทั้งยังคอยให้คำปรึกษาในการทำงานจนสำเร็จได้

สุดท้ายนี้ขอขอบพระคุณครอบครัว รวมถึงเพื่อน ๆ ที่คอยให้กำลังใจตลอดมาจนโครงงานนี้สำเร็จได้ด้วยดี

นาย ณัฐชัย พิมพ์สวัสดิ์

นางสาว นลินพร ศานติคณาวงศ์

สารบัญ

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญ (ต่อ)

หน้า

สารบัญรูป

รูปที่ **หน้า**

สารบัญรูป (ต่อ)

รูปที่ **หน้า**

สารบัญรูป (ต่อ)

รูปที่ **หน้า**

บทที่ 1
บทนำ

1.1 ความสำคัญและที่มาของโครงงาน

เนื่องจากระบบงานส่วนใหญ่ในปัจจุบันมีการนำระบบงานด้านเทคโนโลยีสารสนเทศเข้ามาเกี่ยวข้องกับระบบงานด้วย ซึ่งระบบเทคโนโลยีสารสนเทศที่นำมาใช้งานมักจะเป็นระบบที่มีขนาดใหญ่ ส่งผลให้เกิดความซับซ้อนในการบริหารจัดการ และการดูแลทำได้ยากขึ้น โดยเฉพาะการตรวจสอบว่าการทำงานใดที่ทำให้มีการใช้ทรัพยากรของระบบสูงมากจนผิดปกติ อีกทั้งตามที่กฎหมายพระราชบัญญัติว่าด้วยการกระทำผิดเกี่ยวกับคอมพิวเตอร์กำหนดให้มีการตรวจสอบและจัดเก็บการกระทำต่าง ๆ ของผู้ใช้งานที่เกิดขึ้นในระบบเพื่อไว้ตรวจสอบการทำงาน และเพื่อที่จะใช้เป็นหลักฐานแกะรอย สืบหา หากเกิดการกระทำความผิด ดังนั้นปัญหาที่เกิดขึ้นก็คือ ในระบบงานขนาดใหญ่ การเก็บข้อมูลสำหรับนำมาตรวจสอบและค้นหาสาเหตุของปัญหาด้วยตัวเองนั้นมีเป็นจำนวนมาก ก่อให้เกิดความยากในการนำข้อมูลเหล่านั้นมาตรวจสอบและวิเคราะห์หาสาเหตุของปัญหาด้วยตนเอง นอกจากการจัดเก็บรายละเอียดของการทำงานของระบบแล้วยังต้องมีการตรวจสอบและติดตามการใช้งานทรัพยากรของระบบในทุกส่วน เช่น หน่วยประมวลผล หน่วยความจำ เป็นต้น ซึ่งถือเป็นปัจจัยหลักในการให้บริการในระบบไอที ซึ่งการตรวจสอบว่าการทำงานใดที่มีการใช้งานทรัพยากรระบบสูงมากจนผิดปกติ และส่งผลต่อระบบงานโดยรวมยังไม่สามารถทำได้ จำเป็นต้องใช้ซอฟต์แวร์เฉพาะ

จากปัญหาข้างต้นที่กล่าวมาแล้วทั้งหมด ทางกลุ่มผู้วิจัยจึงทำการพัฒนาโปรแกรมสำหรับค้นหาเหตุการณ์ที่เป็นสาเหตุให้มีการใช้งานทรัพยากรในระบบสารสนเทศสูงมากจนผิดปกติ โดยการรวบรวมข้อมูลการทำงานของระบบสารสนเทศทั้งหมด และข้อมูลการใช้งานทรัพยากรของอุปกรณ์ต่าง ๆ ภายในระบบมาเก็บไว้ที่แหล่งเดียวกันแล้วนำมาทำการวิเคราะห์โดยใช้เทคนิคทางสถิติ เพื่อดูความสัมพันธ์ของข้อมูลระหว่างการทำงานและการใช้ทรัพยากรของระบบ ว่าถ้ามีค่าความเชื่อมั่น (Confidence) ของกฎความสัมพันธ์ที่ได้จากผลวิเคราะห์และค่าการใช้งานทรัพยากรระบบเกิดขึ้นในลักษณะนี้ แล้วค่าดังกล่าวจะส่งผลต่อการทำงานของระบบในลักษณะใด ซึ่งผลที่ได้ทั้งหมดถูกรวบรวมและแสดงผลลัพธ์ออกมาในรูปแบบของเว็บแอพพลิเคชัน (Web Application)

โครงงานนี้จะประกอบด้วยส่วนสำคัญของระบบหลัก ๆ สามส่วน คือ ส่วนแรกเป็นส่วนที่ทำการรวบรวมข้อมูลการทำงานของระบบและรายละเอียดการใช้งานทรัพยากรของระบบโดยทำการจัดเก็บลงในฐานข้อมูลเดียวกัน ส่วนที่สองเป็นการนำข้อมูลจากฐานข้อมูลมาทำการวิเคราะห์ด้วย

โปรแกรมที่ทางกลุ่มวิจัยพัฒนาขึ้น และส่วนที่สามเป็นการนำเสนอผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล เพื่อให้ผู้ดูแลระบบสามารถนำไปใช้ตรวจสอบความผิดปกติที่เกิดขึ้นในระบบสารสนเทศได้สะดวกขึ้น

1.2 วัตถุประสงค์ของโครงงาน

1) เพื่อสร้างระบบที่สามารถจัดเก็บและรวบรวมล็อกไฟล์ (Log File) การทำงานและการใช้ทรัพยากรของระบบไว้ในฐานข้อมูลเดียวกัน ทำให้สะดวกในการนำไปใช้งาน

2) เพื่อสร้างโปรแกรมที่ทำการวิเคราะห์ข้อมูลต่าง ๆ ที่เกิดขึ้นได้อย่างเหมาะสม

3) เพื่อสร้างโปรแกรมที่ทำการวิเคราะห์หาความสัมพันธ์ที่เกิดขึ้นภายในชุดของข้อมูล

4) เพื่อสร้างระบบตรวจสอบความต้องการของแอพพลิเคชัน (Application Request) ที่มีการใช้งานทรัพยากรสูงหรือมีการใช้งานผิดปกติ

5) เพื่อเป็นการช่วยให้ผู้ดูแลระบบสามารถทำการตรวจสอบถึงปัญหาที่เกิดขึ้นในระบบสารสนเทศขนาดใหญ่ได้ง่ายมากขึ้น

1.3 ขอบเขตของโครงงาน

โครงงานนี้ทำการศึกษาเพื่อสร้างโปรแกรมที่สามารถตรวจสอบการทำงานและการใช้งานทรัพยากรของระบบ โดยทำการรวบรวมรายละเอียดการทำงานและการใช้งานทรัพยากรของระบบและนำไปจัดเก็บลงในฐานข้อมูลเดียวกัน และสร้างโปรแกรมที่ทำการนำข้อมูลดังกล่าวมาวิเคราะห์หาความสัมพันธ์ด้วยเครื่องมือการวิเคราะห์ธุรกิจอย่างชาญฉลาด โดยใช้การทำเหมืองข้อมูล (Data Mining) ในเรื่องกฎความสัมพันธ์ (Association Rule) มาเป็นอัลกอริทึม (Algorithm) ของโปรแกรมที่ใช้ในการทำการวิเคราะห์ข้อมูล แล้วนำผลลัพธ์ที่ได้นำเสนอเป็นกราฟของความสัมพันธ์ระหว่างค่าความเชื่อมั่นของกฎความสัมพันธ์ และค่าการใช้งานทรัพยากรระบบในรูปแบบของเว็บแอพพลิเคชัน

1.4 วิธีการดำเนินการ

1) ทำการศึกษาเกี่ยวกับวิธีการรวบรวมล็อกไฟล์จากระบบสารสนเทศจำลอง
2) ทำการศึกษาเกี่ยวกับวิธีการทำเหมืองข้อมูล และเทคนิคการวิเคราะห์ข้อมูลในรูปแบบต่าง ๆ
3) ทดลองใช้งานโปรแกรม WEKA เพื่อศึกษาการทำงานแต่ละฟังก์ชัน (Function)
4) ทำการติดตั้งเพื่อจำลองระบบลงในโปรแกรมจำลองการทำงาน เช่น โปรแกรม Virtual Box
5) ทำการศึกษาและทดลองใช้ฟังก์ชันพื้นฐานของซิสล็อก (Syslog)

6) ทำการรวบรวมรายละเอียดการทำงานและการใช้งานทรัพยากรของระบบทั้งหมด และจัดเก็บข้อมูลทั้งหมดลงไปยังฐานข้อมูลกลาง
7) นำทฤษฎีต่าง ๆ ที่ได้ศึกษามาประยุกต์เพื่อพัฒนาโปรแกรมที่ใช้ทำการวิเคราะห์ข้อมูล
8) นำชุดข้อมูลทดสอบกับโปรแกรมที่ได้พัฒนาขึ้น เพื่อดูผลลัพธ์ที่ได้
9) นำข้อมูลที่ได้จัดเก็บไว้ในฐานข้อมูลเรียบร้อยแล้วมาทำการวิเคราะห์ด้วยโปรแกรมที่พัฒนาขึ้น
10) นำเสนอผลลัพธ์ที่ได้จากการวิเคราะห์ผ่านทางเว็บแอพพลิเคชัน พร้อมทั้งสรุปผลของการทดลองนี้
11) จัดเตรียมการทำรายงานและเตรียมการนำเสนอโครงงาน
12) นำเสนอโครงงาน

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1) ได้รับความรู้ความเข้าใจเกี่ยวกับการติดตั้งและการเรียกใช้งานซิสล็อก
2) ได้รับความรู้ความเข้าใจเกี่ยวกับวิธีการวิเคราะห์หาความสัมพันธ์ของชุดข้อมูลขนาดใหญ่
3) ช่วยให้ผู้ดูแลระบบสามารถตรวจสอบ วิเคราะห์ปัญหาของระบบได้ง่ายและรวดเร็วขึ้น

1.6 ส่วนประกอบของรายงาน

รายงานฉบับนี้ได้แบ่งเนื้อหาออกเป็น 5 บทด้วยกัน คือ

บทที่ 1 บทนำ กล่าวถึงความสำคัญและที่มาของโครงงาน วัตถุประสงค์ของโครงงาน ขอบเขตของโครงงาน วิธีการดำเนินการ ประโยชน์ที่คาดว่าจะได้รับ และส่วนประกอบของรายงาน

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง กล่าวถึงทฤษฎีพื้นฐานที่ใช้ในการทำการรวบรวมข้อมูลรายละเอียดในระบบ และการทำเหมืองข้อมูล

บทที่ 3 การออกแบบและพัฒนา กล่าวถึงภาพรวมของระบบ การออกแบบตารางเก็บข้อมูลต่าง ๆ ส่วนการรวบรวมข้อมูล ส่วนการวิเคราะห์ข้อมูล ส่วนการแสดงผล และการตั้งค่าระบบ

บทที่ 4 การทดลองและผลการทดลอง กล่าวถึงผลลัพธ์ที่ได้การจัดเก็บรายละเอียดของระบบและการใช้งานทรัพยากรระบบ รวมถึงผลลัพธ์ที่ได้จากการหาความสัมพันธ์ของข้อมูลด้วยโปรแกรมวิเคราะห์ข้อมูล และการแสดงผลด้วยเว็บแอพพลิเคชัน

บทที่ 5 บทสรุป กล่าวถึงบทสรุปของโครงงาน ปัญหาอุปสรรค แนวทางการแก้ไข และแนวทางในการพัฒนาโครงงานต่อ

บทที่ 2
ทฤษฎีที่เกี่ยวข้อง

การศึกษาทฤษฎีในโครงงานนี้ได้ทำการศึกษาแบ่งออกเป็น 2 ส่วนหลัก คือ ส่วนแรกทำการศึกษาวิธีการจัดเก็บและการรวบรวมล็อกไฟล์การทำงานและการใช้ทรัพยากรของระบบสารสนเทศ และอีกส่วนหนึ่งทำการศึกษาเกี่ยวกับการทำเหมืองข้อมูลด้วยการใช้เทคนิคกฎความสัมพันธ์ (Association Rule)

2.1 การรวบรวมข้อมูลรายละเอียดในระบบ

ปัจจุบันทุก ๆ ระบบต้องมีการจัดเก็บล็อกไฟล์ต่าง ๆ ซึ่งเป็นสิ่งที่บ่งบอกถึงการกระทำที่เกิดขึ้นในระบบและล็อกไฟล์นี้เป็นสิ่งที่ใช้ในการตรวจสอบว่าระบบนั้น ๆ มีปัญหาอะไรและควรแก้ปัญหาอย่างไรหากไม่มีการเก็บล็อกไฟล์ไว้ใช้ในการตรวจสอบจะทำให้ไม่สามารถทราบว่าปัญหาต่าง ๆ ที่เกิดขึ้นในระบบนั้นเกิดจากสาเหตุใด

ด้วยเหตุนี้ทุกระบบจึงต้องมีการจัดเก็บล็อกไฟล์ของแต่ละเซิร์ฟเวอร์ไว้และเมื่อมีการกระทำเช่นนี้เกิดขึ้น ปัญหาต่อมาที่เกิดขึ้นคือหากระบบที่ใช้เป็นระบบขนาดใหญ่เมื่อเกิดปัญหาหรือต้องการตรวจสอบล็อกไฟล์นั้นเป็นสิ่งที่ทำได้ยากลำบาก รวมถึงความลำบากในการจัดการล็อกไฟล์ในแต่ละเซิร์ฟเวอร์ด้วย เนื่องจากปัญหาดังที่กล่าวมาแล้ว จึงได้มีการนำโปรแกรมที่ชื่อว่าซิสล็อก (Syslog) มาใช้ในการแก้ปัญหา โดยโปรแกรมนี้จะทำการเก็บรวบรวมล็อกไฟล์จากแหล่งต่าง ๆ มาจัดเก็บที่เซิร์ฟเวอร์ส่วนกลาง ทำให้ล็อกจากทุก ๆ แหล่งถูกจัดเก็บมาไว้ในที่เดียวกันเพื่อให้สะดวกต่อการจัดการล็อกไฟล์และการตรวจสอบสิ่งที่เกิดขึ้นในระบบอีกด้วย

2.1.1 ซิสล็อก

ซิสล็อกเป็นกลไกใช้ในการเก็บข้อมูลล็อกของเคอร์เนล (Kernel) และแอพพลิเคชันบนระบบปฏิบัติการยูนิกซ์และลินุกซ์ที่ถูกติดตั้งมาให้พร้อมกับระบบปฏิบัติการในเกือบทุกระบบโดยผู้ดูแลระบบสามารถปรับแต่งไฟล์การตั้งค่า (Configuration) เพื่อควบคุมการทำงานของซิสล็อกได้ เช่น ให้ซิสล็อกเก็บข้อมูลไปไว้ที่ไฟล์ใดหรือให้ส่งข้อมูลล็อกนี้ไปเก็บไว้ยังเครื่องอื่นในเครือข่าย ข้อมูลล็อกที่ควบคุมโดยซิสล็อกนั้นจะถูกกำหนดให้มีค่า facility และ priority โดยส่วนของ facility นั้นเป็นข้อมูลที่อธิบายถึงแหล่งกำเนิดของข้อมูลล็อกนั้น ๆ เช่น ข้อมูลล็อกที่ส่งมาจากระบบเมล์ก็จะมี

facility เป็นเมล์ (Mail) ส่วน priority นั้นจะแสดงถึงระดับความสำคัญของเหตุการณ์ที่เกิดสำหรับแต่ละ facility ทั้งนี้ข้อมูลล็อกทุกอันจำเป็นต้องมี facility และ priority เสมอ

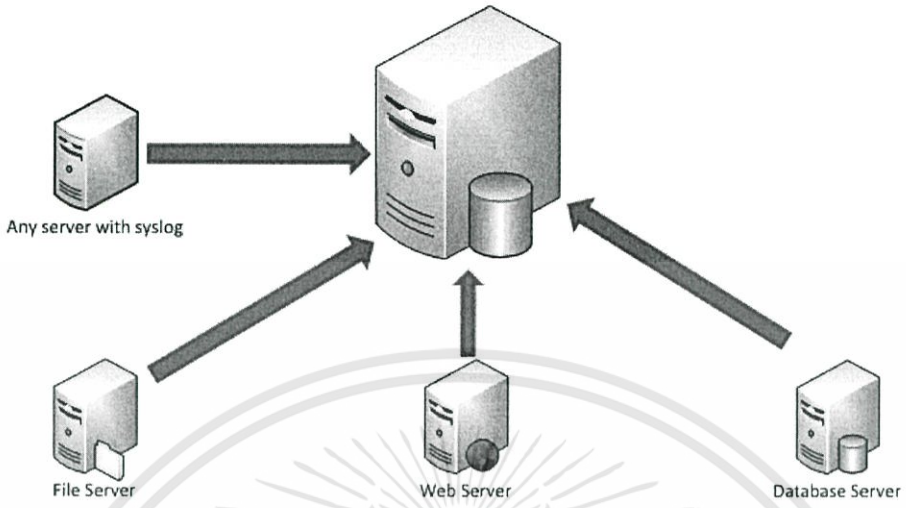
ซิสล็อกนั้นเป็นมาตรฐานสำหรับการทำล็อกกิ้ง (Logging) ของยูนิกซ์และลินุกซ์แต่ในปัจจุบันซิสล็อกกำลังจะถูกแทนที่ด้วยซิสล็อก-เอ็นจี (Syslog-NG) ซึ่งมีความยืดหยุ่นมากกว่าซิสล็อกและสามารถเก็บข้อมูลล็อกบนพื้นฐานของนิพจน์ปกติ (Regular Expression) ได้ ซึ่งข้อเสียของซิสล็อกมีดังต่อไปนี้

1) เนื่องจากซิสล็อกเป็นการส่งข้อมูลแบบยูดีพี (UDP) ทำให้ผู้ส่งไม่ได้รับความมั่นใจว่าข้อมูลที่ส่งไปให้เครื่องเซิร์ฟเวอร์นั้นจะไปถึงหรือไม่
2) การที่ไม่มีการระบุตัวตน (Authentication) เพื่อยืนยันว่าใครเป็นผู้ส่งนำไปสู่การส่งข้อมูลแปลกปลอมปนเข้าไปยังเครื่องล็อกเซิร์ฟเวอร์หากผลเป็นเช่นนี้อาจส่งผลให้การติดตามรอยของผู้บุกรุกเกิดการผิดพลาดได้
3) การส่งข้อมูลข้อความต้นฉบับ (Plaintext) โดยไม่มีการเข้ารหัสก่อนที่จะส่งอาจจะทำให้ผู้ไม่ประสงค์ดีสามารถดักจับข้อมูลที่สำคัญไปได้และอาจจะทาให้ผู้บุกรุกรู้ได้ว่าระบบของเรานั้นมีช่องโหว่ที่จะสามารถนาไปใช้ประโยชน์อย่างไรได้บ้าง

ในระบบยูนิกซ์ส่วนใหญ่จะยังคุ้นเคยกับซิสล็อกเป็นอย่างดี เพราะซิสล็อกถือได้ว่าเป็นซิสล็อก-ดีลมอลที่ใช้กันมาอย่างยาวนานและกลายเป็นมาตรฐานของการเก็บข้อมูลล็อกของระบบปฏิบัติการยูนิกซ์แต่อย่างไรก็ตามซิสล็อกก็ยังมีข้อเสียบางอย่างดังที่กล่าวมาแล้ว ซึ่งล็อกดีลมอลตัวอื่น เช่น ซิสล็อก-เอ็นจี, เอ็มซิสล็อก สามารถแก้ไขข้อบกพร่องดังกล่าวได้ ดังนี้

1) ซิสล็อก-เอ็นจี สามารถทำงานได้ทั้งบนทีซีพี (TCP) และยูดีพี (UDP)
2) ซิสล็อก-เอ็นจี สามารถทำการกรอง (Filter) ข้อมูลได้ด้วยนิพจน์ปกติ
3) ซิสล็อก-เอ็นจี สามารถทำงานในรูปแบบที่อ้างอิง priority/facility ได้ดังนั้นจึงสามารถทำงานแทนซิสล็อกได้
4) ซิสล็อก-เอ็นจี สนับสนุนล็อกฟอเวิดดิง (Log Forwarding) ซึ่งทำให้สามารถทราบได้ว่าต้นทางของล็อกถูกส่งมาจากเครื่องใดและผ่านเครื่องใดมาบ้าง

นอกจากนี้ซิสล็อก-เอ็นจียังมีรูปแบบของไฟล์การตั้งค่าที่ง่ายแต่มีความยืดหยุ่นสูง สามารถนำไปประยุกต์ใช้ให้ตรงตามความต้องการได้โดยง่ายและควรรันซิสล็อก-เอ็นจีภายหลังจากการสร้างไฟล์การตั้งค่าเสร็จสิ้นแล้วเท่านั้น โดยซิสล็อก-เอ็นจีนี้มีการเลือก (Option) ในการรันค่อนข้างง่าย

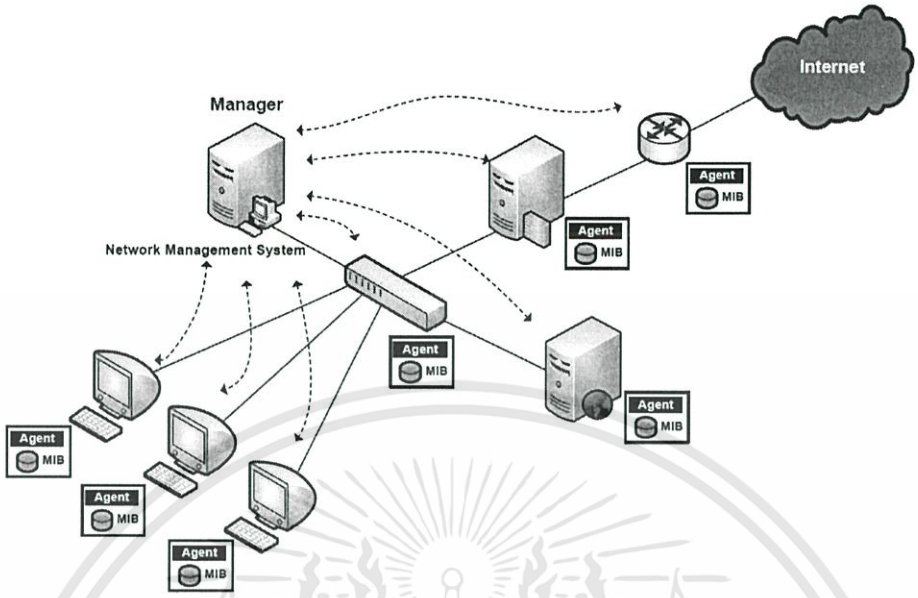


รูปที่ 2.1 การจัดการระบบโดยใช้ซิสล็อก

จากรูปที่ 2.1 เป็นการทำงานโดยรวมของระบบโดยใช้ซิสล็อกเป็นโปรแกรมที่ช่วยในการรับส่ง ล็อกไฟล์จากเซิร์ฟเวอร์ต่าง ๆ ในระบบไปสู่เซิร์ฟเวอร์ที่เป็นส่วนรวบรวมล็อกไฟล์โดยหลักการทำงานนั้นจะเริ่มขึ้นเมื่อเซิร์ฟเวอร์ต่าง ๆ ในระบบมีการทำงานโดยเกิดจากเรียกใช้งานของผู้ใช้ เซิร์ฟเวอร์เหล่านั้นทำให้เกิดล็อกไฟล์ขึ้นและเมื่อโปรแกรมซิสล็อกพบว่ามีล็อกไฟล์ใหม่เข้ามาในระบบ โปรแกรมจึงทำการนำล็อกไฟล์เหล่านั้นส่งต่อไปยังส่วนรวบรวมล็อกไฟล์ทันทีโปรแกรมนี้จะทำงานแบบนี้ตลอดเวลาเมื่อมีการเริ่มต้นทำงานของโปรแกรม

2.1.2 เอสเอ็นเอ็มพี (SNMP : Simple Network Management Protocol)

โปรโตคอลเอสเอ็นเอ็มพี เป็นโปรโตคอลประยุกต์ที่กำหนดรูปแบบและกรรมวิธีในการบริหารจัดการเครือข่าย โดยจะมีสถานีจัดการเครือข่ายส่วนกลางเรียกว่า เอสเอ็นเอ็มพีเอเจนต์ (SNMP Agent) ทำหน้าที่ติดต่อประสานงานระหว่างเอ็นเอ็มเอส (NMS) กับอุปกรณ์เครือข่ายที่เป็นเอเจนต์ต่าง ๆ เช่น พีซี (PC : Personal Computer), โมเด็ม (Modem), ฮับ (Hub), สวิตช์ (Switch) เราเตอร์ (Router) เป็นต้น โดยเอสเอ็นเอ็มพีเอเจนต์นี้จะเชื่อมต่ออยู่กับส่วนทำงานที่เป็นฮาร์ดแวร์ (Hardware) และซอฟต์แวร์ (Software) ของเอเจนต์ ซึ่งจะมีหน้าที่ในการคอยรับคำสั่งในการปรับเปลี่ยนการทำงานของเอเจนต์จากเอ็นเอ็มเอสและคอยรายงานข้อมูลของเอเจนต์เมื่อเอ็นเอ็มเอสร้องขอ โดยจะมีการทำการยืนยันสิทธิการร้องขอข้อมูลและปรับเปลี่ยนค่าของเอ็นเอ็มเอสในรูปรหัสผ่านสำหรับการเชื่อมต่อระหว่างเอสเอ็นเอ็มพีเอเจนต์กับส่วนที่ทำงานเป็นฮาร์ดแวร์และซอฟต์แวร์ของเอเจนต์แสดงได้ดังรูปที่ 2.2



รูปที่ 2.2 แสดงการเชื่อมต่อเอสเอ็นเอ็มพีเอเจนต์

การติดต่อระหว่างเอ็นเอ็มเอสกับเอเจนต์มีรูปแบบในการติดต่อมากมายขึ้นอยู่กับวัตถุประสงค์ในการติดต่อแต่สาหรับเอสเอ็นเอ็มพีรุ่น 1 (SNMP Version 1) มีรูปแบบในการติดต่อกัน 5 แบบ ดังนี้

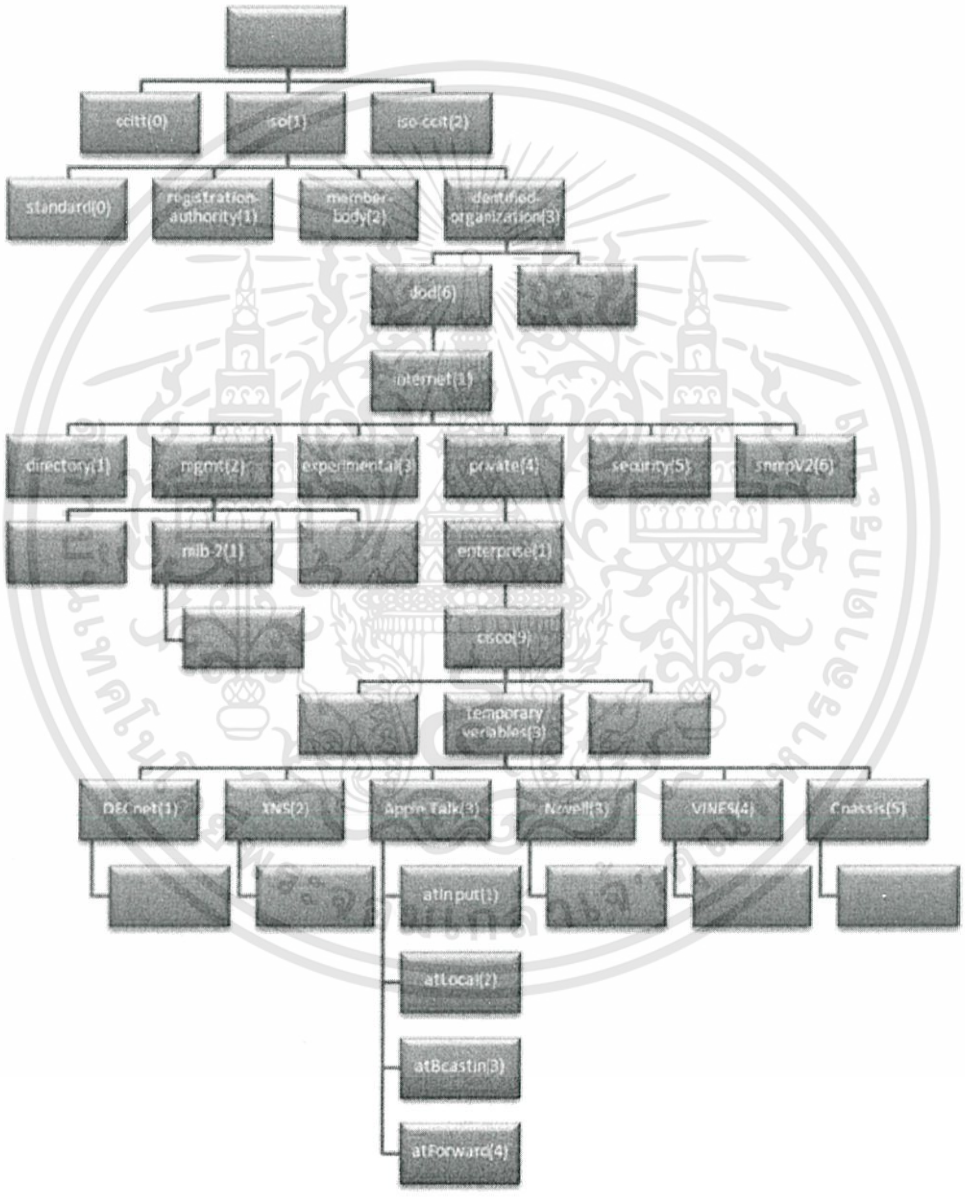
1) get-request คือ ใช้สอบถามข้อมูลจากเอเจนต์ที่อยู่บนอุปกรณ์ที่ต้องการตรวจสอบในเครือข่าย
2) get-next-request คือ ใช้สอบถามข้อมูลที่เรียงเป็นลำดับเช่นข้อมูลที่เก็บอยู่ในรูปแบบตารางหรือในกรณีไม่ทราบชื่อตัวแปรที่แน่ชัด
3) get-response คือ เอเจนต์ส่งคำตอบกลับมายังผู้สอบถาม
4) set-request คือ ใช้เปลี่ยนแปลงค่าตัวแปรที่เอเจนต์รับผิดชอบอยู่
5) trap คือ ใช้แจ้งเหตุการณ์ที่เกิดขึ้นในระบบเครือข่ายเช่น การเริ่มต้นทำงานใหม่ของอุปกรณ์หรือเส้นทางที่ขัดข้องเอสเอ็นเอ็มพีอาศัยโปรโตคอลยูดีพีในการติดต่อซึ่งต้องอาศัยพอร์ต (Port) ต่าง ๆ ในการติดต่อโดยเอสเอ็นเอ็มพีใช้พอร์ตหมายเลข 161 ในการติดต่อ

- เอ็มไอบี (MIB : Management Information Base)

เอ็มไอบีหรือฐานข้อมูลสารสนเทศจัดการมีหน้าที่ในการเก็บตัวแปรและค่ากำหนดการทำงานประจำอุปกรณ์ซึ่งข้อมูลประจำอุปกรณ์แต่ละตัวสามารถมีได้หลากหลายและอุปกรณ์ต่างชนิดกันก็

ย่อมมีข้อมูลประจำอุปกรณ์ต่างกันอีกด้วย ดังนั้นการสอบถาม (เพื่ออ่านค่า) หรือการเปลี่ยนแปลงค่า (เพื่อเขียนค่า) ในฐานข้อมูลจึงต้องมีรูปแบบที่เป็นมาตรฐานให้กับอุปกรณ์ทุกประเภท

โครงสร้างที่เหมาะสมที่สุดสำหรับใช้เป็นฐานข้อมูลเพื่อจัดเก็บตัวแปรเหล่านี้คือโครงสร้างต้นไม้แบบลำดับชั้น (Hierarchy Tree) เราเรียกโครงสร้างต้นไม้ของฐานข้อมูลนี้ว่าเอ็มไอบีทรี (MIB Tree) ซึ่งแสดงได้ดังรูปที่ 2.3



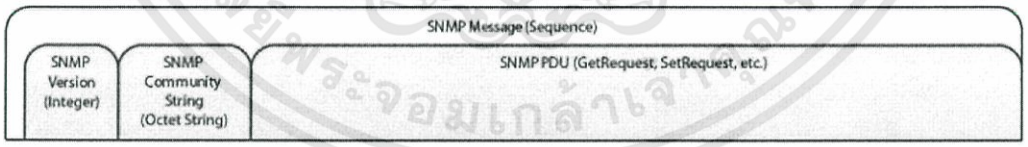
รูปที่ 2.3 เอ็มไอบีทรี

จากรูปที่ 2.3 ในแต่ละโหนด (Node) ของเอ็มไอบีทรีนี้จะใช้แทนวัตถุ (Object) ที่มีชื่อพร้อมทั้งเลขฐานสิบกำกับอยู่ประจำโหนดเพื่อใช้ในการอ้างอิงยกเว้นโหนดราก (RootNode) ที่จะไม่มีชื่อกำกับ

ในระดับแรกของเอ็มไอบีทรีจะมีโหนดหลัก 3 โหนดซึ่งกำหนดกลุ่มองค์กร 3 กลุ่ม คือ ITUT(0), ISO(1) และ Joint-ISO-ITU-T(2) ภายใต้โหนด ISO มีโหนดลำดับที่ 3 คือ org(3) กำหนดองค์กรนานาชาติและส่วนหนึ่งขององค์กรนี้ คือ dod(6) หรือ Department of Defense และมีโหนด internet(1) เพื่อกำหนดกลุ่มการจัดการเครือข่ายในอินเทอร์เน็ต เมื่อต้องการอ้างอิงถึงโหนดในโครงสร้างให้เขียนหมายเลขจากรากไปตามเส้นทางถึงโหนดนั้นและคั่นด้วยจุด เราจะเรียกลำดับตัวเลขนี้ว่าโอไอดี (OID : Object Identifier) เช่น 1.3.6.1.2.1.1 เป็นโอไอดีที่มีค่าเท่ากับชื่อ iso.org.dod.internet.mgmt.mib-2.system โหนดที่อยู่ภายใต้ 1.3.6.1.2.1 หรือในกลุ่ม mib-2 เป็นโหนดสำหรับการใช้งานเอสเอ็นเอ็มพีแต่ละโหนดจะมีโหนดย่อยเพื่ออ้างอิงถึงตัวแปรต่าง ๆ เช่น 1.3.6.1.2.1.1.1 คือ ตัวแปร sysDescr (System Description) ซึ่งทำหน้าที่เก็บคำอธิบายเกี่ยวกับอุปกรณ์นั้น

- รูปแบบข้อความของเอสเอ็นเอ็มพี

รูปแบบ (Format) ของข้อความ (Massage) SNMP จะระบุว่าช่องใน SNMP อยู่ตำแหน่งใดและเรียงลำดับอย่างไร ยิ่งกว่านั้นข้อความอาจจะมีหลายชั้นของช่องที่ซ้อนกัน ณ ที่ชั้นนอกสุดข้อความของ SNMP เป็นช่องเดี่ยวประเภทลำดับ (Sequenc) ภาพรวมของข้อความทั้งหมดจะเป็นแถวของช่อง 3 ช่องนั้นคือ ช่องเวอร์ชัน (Version) ของ SNMP บรรจุค่าเป็นอินทิเจอร์ (Integer), ช่องคอมมิวนิตี้สตริง (Community String) บรรจุค่าเป็นออกเทตสตริง (Octet String) และ พีดียู (PDU) บรรจุค่าของเก็ต รีเควสและเซ็ต รีเควส (Get Request, Set Request)

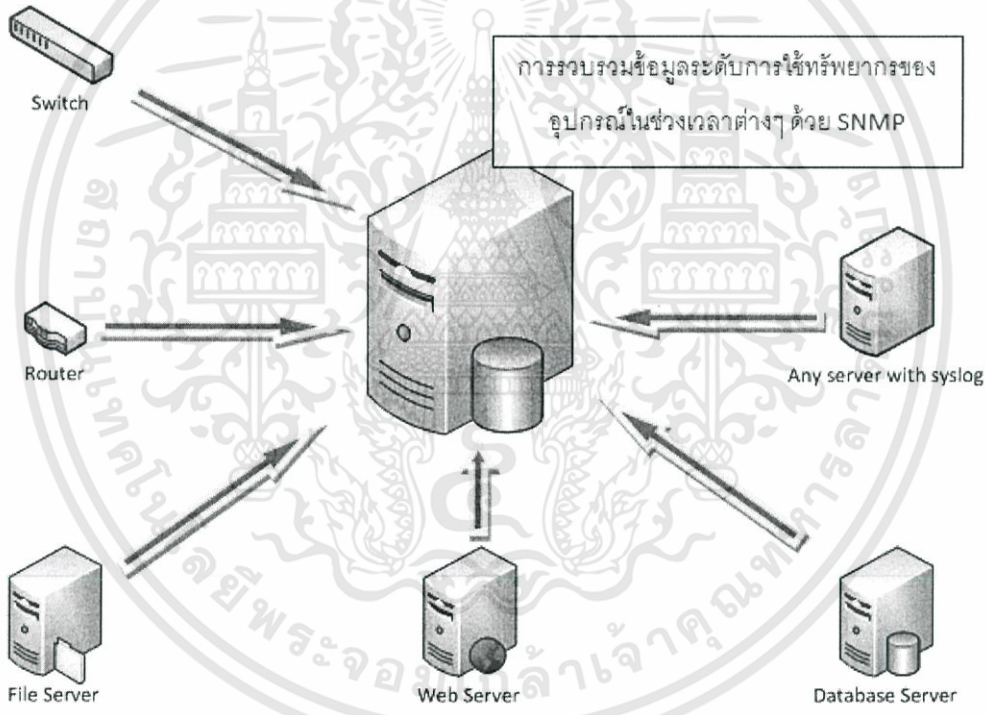


รูปที่ 2.4 ฟอร์แมตพื้นฐานของเอสเอ็นเอ็มพี

ช่องเวอร์ชันและช่องคอมมิวนิตี้สตริงจะเป็นชนิดข้อมูลแบบไพรมิทีฟ (Primitive) ที่ไม่ได้สร้างจากช่องเล็ก ๆ หลายช่องหรือมีหลายชั้นแต่สำหรับช่องพีดียู จะเป็นชนิดข้อมูลแบบคอมเพล็กซ์ (Complex) ที่ประกอบด้วยหลายช่องขนาดเล็ก (หลายเลเยอร์) ภายในพีดียู จะประกอบด้วยช่อง

รีเควสไอดี (Request ID), ช่องเออร์เรอร์ (Error), ช่องเออร์เรอร์อินเดกซ์ (Error Index) และช่องวาร์ไบด์ลิสต์ (Varbind List)

วาร์ไบด์จะประกอบด้วย 2 ช่องที่ต่อกันเป็นคู่ ๆ กันไป ช่องแรกคือ ค่าโอไอดี (OID) ที่ระบุที่อยู่ของพารามิเตอร์ที่ต้องการ ช่องที่สองคือ ช่องแวลู (Value) คือ ค่าของพารามิเตอร์ที่ต้องการสำหรับคำสั่งเซ็ตรีเควส ค่าที่ตั้งค่าต้องเป็นชนิดข้อมูลเดียวกันกับชนิดข้อมูลที่ระบุในเอ็มไอบี (MIB) ของอุปกรณ์นั้น ๆ สำหรับคำสั่งเก็ตรีเควส ช่องแวลูจะมีค่าเท่ากับนัล (Null) ด้วยค่าความยาวเท่ากับ 0 x 00 ค่านัลนี้ก็เป็นตัวบ่งบอกตัวเอเจนต์ให้ส่งค่ากลับด้วยเก็ตรีเควส พีดียู สำหรับวาร์ไบด์ลิสต์คือแถวลำดับของวาร์ไบด์ในกรณีข้อความตั้งค่าหรือดึงค่าเพียงหนึ่งพารามิเตอร์วาร์ไบด์ลิสต์จะมีแค่หนึ่งวาร์ไบด์



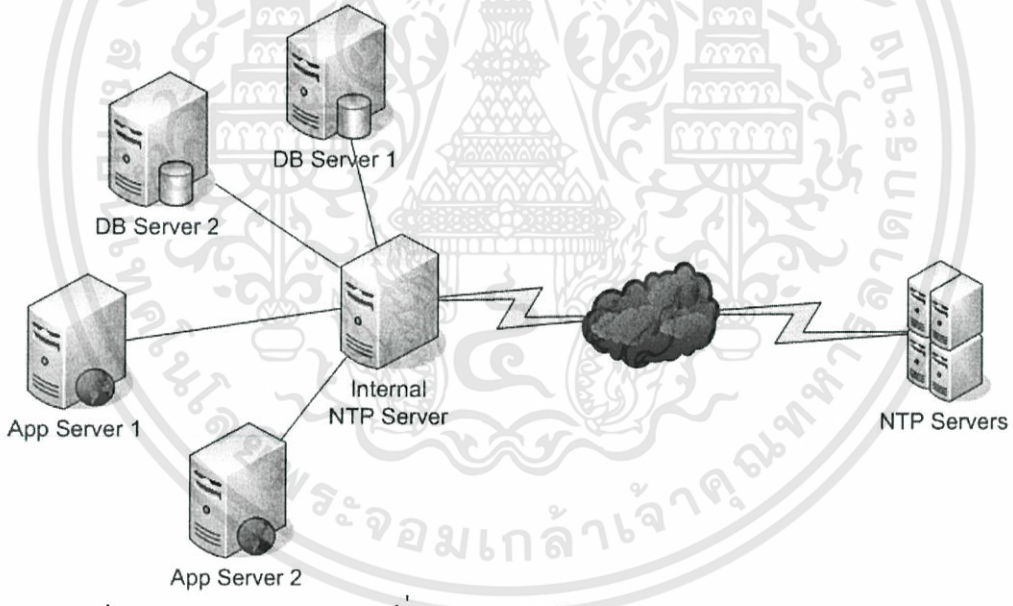
รูปที่ 2.5 โครงสร้างระบบที่บริหารจัดการด้วยเอสเอ็นเอ็มพี

จากรูปที่ 2.5 หลักการทำงานของเอสเอ็นเอ็มพีเซิร์ฟเวอร์ (SNMP Server) ส่วนกลาง จะมีการดึงข้อมูลการใช้งานทรัพยากรจากเซิร์ฟเวอร์ (Server) ต่าง ๆ ในระบบมาจัดเก็บลงฐานข้อมูลส่วนกลางเพื่อให้ง่ายต่อการจัดการข้อมูลผ่านทางโปรโตคอลเอสเอ็นเอ็มพี ซึ่งในการดึงข้อมูลนั้นจะทำการดึงมา 3 อย่าง คือ การใช้งานซีพียู (CPU), การใช้งาน (Memory) และการใช้งานแบนด์วิดท์

(Bandwidth) สาเหตุที่ดึง 3 ค่านี้มาทำการวิเคราะห์ เนื่องจากเป็นค่าที่บ่งบอกถึงการใช้งานเซิร์ฟเวอร์ได้อย่างชัดเจนสามารถทำการประเมินการทำงานของเซิร์ฟเวอร์ได้ดีและชัดเจนที่สุด โดยในการดึงค่านั้นสามารถดึงได้จากทุกอุปกรณ์ที่ทำการเชื่อมต่อกับระบบ เช่น เราเตอร์ (Router) สวิตช์ (Switch) เป็นต้น

2.1.3 เวลาของเซิร์ฟเวอร์ในระบบ

ในการจำลองระบบนั้นมีความจำเป็นที่ต้องให้เวลาของระบบเป็นเวลาเดียวกันทั้งหมด เพื่อให้เกิดความถูกต้องของการจัดการเก็บและตรวจสอบล็อกไฟล์ในระบบ จึงมีการออกแบบโดยใช้โปรโตคอลเอ็นทีพี (NTP : Network Time Protocol) เพื่อใช้ในการปรับเทียบเวลาของระบบ โดยให้เครื่องเซิร์ฟเวอร์ส่วนกลางมีการปรับเทียบเวลาจริงกับเซิร์ฟเวอร์ภายนอกซึ่งใช้เวลาสากล จากนั้นจึงเปิดบริการให้เซิร์ฟเวอร์ภายในระบบมีการปรับเทียบเวลากับเซิร์ฟเวอร์ส่วนกลาง เพื่อให้ทุก ๆ เซิร์ฟเวอร์ในระบบมีเวลาที่เท่ากันทั้งหมด จุดประสงค์เพื่อไม่ให้เกิดการเหลื่อมล้ำกันของเวลาเมื่อทำการจัดเก็บข้อมูลลงฐานข้อมูล ซึ่งตัวอย่างแสดงการทำงานดังรูปที่ 2.6



รูปที่ 2.6 โครงสร้างของระบบที่ใช้โปรโตคอลเอ็นทีพี (NTP) ในการปรับเทียบเวลา

2.2 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล เป็นกระบวนการที่กระทำกับข้อมูลซึ่งมีจำนวนมากเพื่อค้นหารูปแบบแนวทาง และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้นโดยอาศัยหลักสถิติ การรู้จำ การเรียนรู้ของเครื่อง

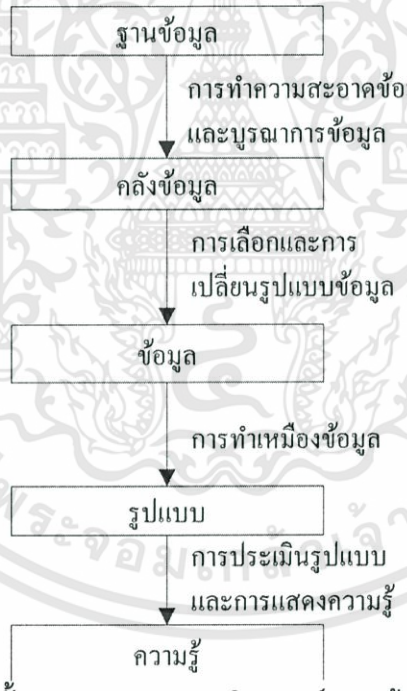
และหลักคณิตศาสตร์ ซึ่งความสัมพันธ์เหล่านี้จะแสดงให้เห็นถึงความรู้ที่เป็นประโยชน์และน่าสนใจ สามารถนำมาประกอบการตัดสินใจได้

รูปที่ 2.7 ขั้นตอนการทำเหมืองข้อมูล

จากรูปที่ 2.7 แสดงขั้นตอนในการทำเหมืองข้อมูล โดยเป็นการนำข้อมูลมาทำการวิเคราะห์หาความรู้และความสัมพันธ์ของข้อมูล เพื่อใช้ในการตัดสินใจและนำข้อมูลนั้นไปใช้งานต่อไป

2.2.1 กระบวนการค้นหาความรู้จากฐานข้อมูล

กระบวนการในการวิเคราะห์ความรู้ในฐานข้อมูลนั้นประกอบไปด้วย 7 ขั้นตอน ดังรูปที่ 2.8



รูปที่ 2.8 ขั้นตอนการะบวนการวิเคราะห์ความรู้จากฐานข้อมูล

1) การทำความสะอาดข้อมูล (Data Cleaning) จุดประสงค์ของการทำความสะอาดข้อมูลนั้นเพื่อทำให้มั่นใจในคุณภาพของข้อมูลที่เลือกมา ว่าคุณภาพของข้อมูลนั้นเหมาะสมแก่การนำไปทำเหมืองข้อมูล สาเหตุที่ต้องทำความสะอาดข้อมูลเพราะว่าข้อมูลที่มีอยู่อาจ

ไม่ครบถ้วนและไม่สมบูรณ์เพียงพอแก่การนำมาวิเคราะห์เพื่อหาความสัมพันธ์ เช่น ข้อมูลขาดหายไป, ข้อมูลไม่สัมพันธ์กัน เป็นต้น และวิธีการทำความสะอาดข้อมูลนั้นมีหลากหลายวิธี เช่น การเติมข้อมูลที่ขาดหายไป, การระบุข้อมูลที่ผิดพลาดหรือข้อมูลที่โดดออกมาจากกลุ่ม เป็นต้น

2) การบูรณาการข้อมูล (Data Integration) จุดประสงค์ของการบูรณาการข้อมูลเพื่อยุบรวมข้อมูลจากหลายแหล่งที่มีรูปแบบการเก็บของข้อมูลที่คล้ายคลึงกันให้เป็นข้อมูลที่มีรูปแบบเดียวกัน เพื่อลดการเกิดความซ้ำซ้อนกันของข้อมูล และเก็บไว้ในที่เดียวกัน

3) การเลือกข้อมูล (Data Selection) เป็นการระบุเฉพาะข้อมูลที่ต้องการ และกำจัดข้อมูลที่ไม่ต้องการ โดยจุดประสงค์เพื่อลดจำนวนของข้อมูลที่ไม่จำเป็นลง ซึ่งจะเป็นการลดภาระในขั้นตอนการทำเหมืองข้อมูล

4) การเปลี่ยนรูปแบบข้อมูล (Data Transformation) จุดประสงค์ในการเปลี่ยนแปลงรูปแบบของข้อมูลเพื่อทำการปรับรูปแบบข้อมูลให้อยู่ในรูปแบบที่เหมาะสมแก่การทำเหมืองข้อมูล วิธีในการเปลี่ยนรูปแบบข้อมูลได้แก่ การสร้างรูปแบบปกติ (Normalization), การสร้างหัวตารางเพิ่ม (Attribute Construction) เป็นต้น

5) การทำเหมืองข้อมูล (Data Mining) จุดประสงค์ในการทำเหมืองข้อมูลเพื่อวิเคราะห์หาความรู้ความสัมพันธ์ของข้อมูล และรูปแบบที่เป็นไปได้ทั้งหมดของข้อมูล ซึ่งขึ้นอยู่กับอัลกอริทึมที่ใช้แต่ละอัลกอริทึมจะมีประสิทธิภาพในการวิเคราะห์ข้อมูลที่แตกต่างกัน

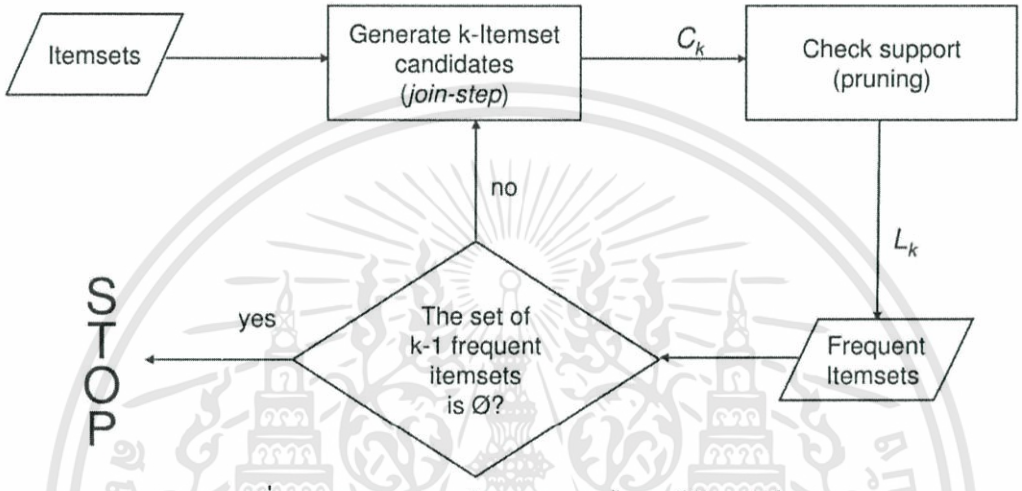
6) การประเมินรูปแบบ (Pattern Evaluation) จุดประสงค์ในการประเมินรูปแบบเพื่อนำความสัมพันธ์และรูปแบบที่ได้จากการทำเหมืองข้อมูลมาวิเคราะห์ตามหลักการทางธุรกิจเพื่อคัดเลือกเฉพาะรูปแบบที่สนใจ ซึ่งหลังจากการวิเคราะห์แล้วจะนำเสนอในรูปแบบอื่น เช่น แสดงด้วยกราฟ เป็นต้น

7) การแสดงความรู้ (Knowledge Presentation) เป็นการนำความรู้ที่ผ่านการวิเคราะห์ตามหลักการทางธุรกิจแล้วมานำเสนอในรูปแบบที่เป็นระเบียบ เข้าใจง่าย เช่น กราฟ, ตาราง และเหมาะสมที่จะนำไปใช้งานได้จริง

2.2.2 กฎความสัมพันธ์ (Association Rule)

กฎความสัมพันธ์เป็นการค้นหาความสัมพันธ์ระหว่างข้อมูลในฐานข้อมูล โดยผลจากการวิเคราะห์จะอยู่ในรูปแบบของกฎความสัมพันธ์ โดยมีค่าสนับสนุน (Support) และค่าความเชื่อมั่น (Confidence) เป็นค่าที่ใช้สำหรับบอกความน่าสนใจของกฎนั้น ๆ โดยอยู่ในรูปแบบของการทำนายที่ว่า หากมีเหตุการณ์ใดเกิดขึ้น แล้วจะมีเหตุการณ์อื่นใดเกิดขึ้นตามมาบ้าง ด้วยความมั่นใจกี่เปอร์เซ็นต์ ซึ่งออไพออรี่อัลกอริทึมถือเป็นอัลกอริทึมพื้นฐานที่ใช้ในการหาความสัมพันธ์ของข้อมูล

โดยใช้หลักการค้นหาแบบวงกว้างก่อน ซึ่งจะทำการสร้างและตรวจสอบกลุ่มข้อมูลที่เกิดขึ้นบ่อยทีละชั้น โดยเริ่มจากกลุ่มข้อมูลที่มีจำนวนสมาชิกเท่ากับหนึ่งถ้ากลุ่มข้อมูลใดมีค่าสนับสนุนน้อยกว่าค่าสนับสนุนที่กำหนดก็จะตัดกลุ่มข้อมูลนั้นออก ไม่นำไปสร้างกลุ่มข้อมูลในชั้นต่อไป การทำงานของอัลกอริทึมจะวนไปเรื่อย ๆ จนกระทั่งไล่ทุกระดับชั้นหรือไม่เหลือกลุ่มของข้อมูลที่จะสร้างในชั้นต่อไป



รูปที่ 2.9 กระบวนการทำงานของออไพออรีอัลกอริทึม

รูปแบบที่เกิดขึ้นบ่อยนั้น คือ รูปแบบหรือชุดของข้อมูลที่เกิดขึ้นบ่อยในฐานข้อมูล ซึ่งมักจะใช้กระบวนการนี้ในฐานข้อมูลแบบทรานแซ็คชัน (Transaction) โดยลักษณะข้อมูลแบบทรานแซ็คชันแสดงได้ดังรูปที่ 2.10

TID	*List of item_IDs*
T100	I1, I2, I5
T200	I2, I4
T300	I2, I3
T400	I1, I2, I4
T500	I1, I3
T600	I2, I3
T700	I1, I3
T800	I1, I2, I3, I5
T900	I1, I2, I3

รูปที่ 2.10 ตัวอย่างข้อมูลแบบทรานแซ็คชัน

ความหมายของทรานแซ็คชันในการทำเหมืองข้อมูลมักจะหมายถึง ข้อมูลในแต่ละแถว (tuple) ซึ่งจะแตกต่างกับความหมายในวิชาฐานข้อมูลที่จะหมายถึง การทำงานเชิงตรรกะ (A logical unit of work) โดยผลลัพธ์ของการหารูปแบบที่เกิดขึ้นบ่อยนั้นจะได้ผลลัพธ์ออกมาในแง่ของกฎแสดงความเชื่อมโยง ซึ่งกฎนี้เป็นกฎที่ใช้แสดงถึงความสัมพันธ์ระหว่างข้อมูลที่เกิดขึ้นพร้อมกัน

การหากลุ่มของข้อมูลที่เกิดขึ้นบ่อยทั้งหมด เป็นขั้นตอนที่จะต้องหาว่าข้อมูลแต่ละกลุ่มข้อมูลนั้นเกิดขึ้นมากกว่าจุดที่กำหนดไว้โดยจุดที่กำหนดนี้เรียกว่า ค่าสนับสนุนที่ต่ำที่สุด (Minimum Support Count) ซึ่งค่านี้คือความเป็นไปได้ที่ทรานแซ็คชันหนึ่ง ๆ จะเกิดข้อมูล 2 ชนิดพร้อมกัน ที่ผู้ใช้กำหนดขึ้นและยอมรับได้โดยทั่วไป ค่านี้สามารถเขียนได้ใน 2 ลักษณะ ดังนี้

1) รูปของเปอร์เซ็นต์ (Relative Support) โดยวิธีการหานั้นหาจากจำนวนที่ข้อมูลเกิดขึ้นพร้อมกัน หารด้วยจำนวนของทรานแซ็คชันทั้งหมด
2) รูปแบบที่เป็นจำนวนเต็ม (Absolute Support) โดยวิธีการหานั้นหานับจำนวนที่ข้อมูลเกิดขึ้นพร้อมกันที่มีอยู่ในทรานแซ็คชัน

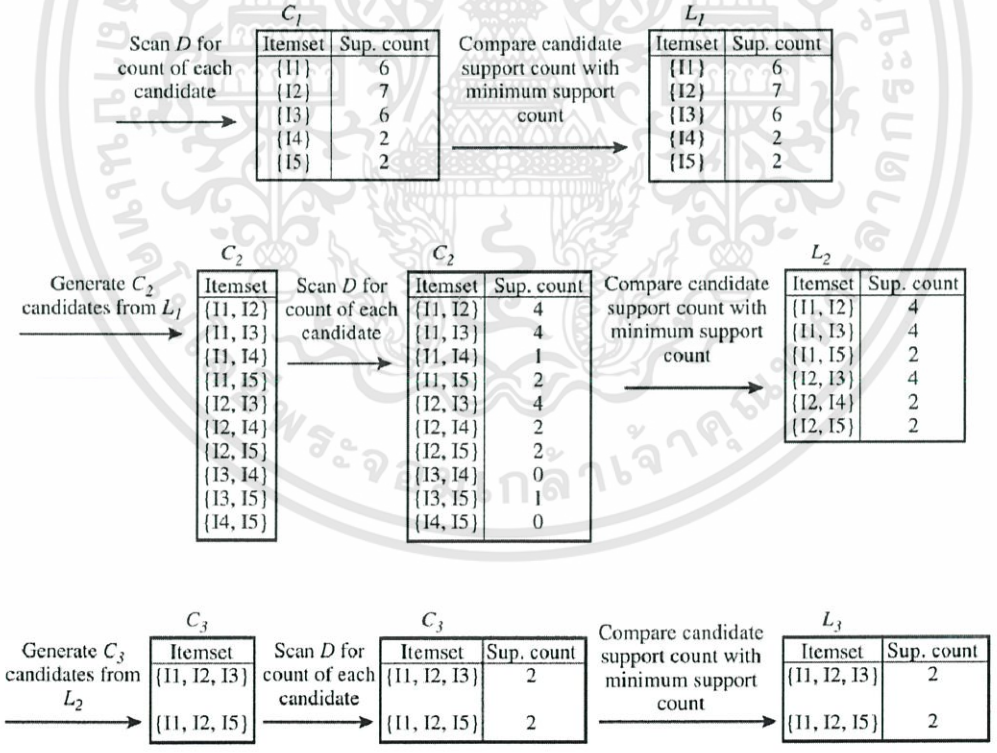
โดยกระบวนการหากลุ่มข้อมูลที่เกิดขึ้นบ่อย (Frequent Itemset) นั้นจะใช้แนวคิดของออไพออรีอัลกอริทึม ซึ่งชื่อของแนวคิดของออไพออรีเป็นชื่อของผู้ที่คิดแนวคิดของการหากลุ่มข้อมูลที่เกิดขึ้นบ่อย โดยแนวคิดของออไพออรี นั้นประกอบด้วย 2 ขั้นตอน คือ

ขั้นตอนแรก คือ การรวมตารางตัวแทนที่ถูกสร้างขึ้นมาเข้ากับตารางตัวแทน โดยขั้นตอนนี้เกิดขึ้นสำหรับสร้างกลุ่มของข้อมูลคู่แข่งขึ้น (A set of candidate itemset or Ck) โดยที่ k คือจำนวนกลุ่มของข้อมูลเช่น B C D ในฐานข้อมูลทรานแซ็คชัน เราจะหมายถึงมี 3 กลุ่มข้อมูลหรือ k=3

และจะใช้ข้อมูลเหล่านี้มาคัดเลือกว่ากลุ่มของข้อมูลใดมีค่ามากว่าค่าสนับสนุนที่ต่ำสุด เราใช้กลุ่มของข้อมูลเหล่านั้นมาเป็นกลุ่มข้อมูลที่เกิดขึ้นบ่อย (A set of frequent itemset or Lk) ซึ่งก่อนจะทำการรวมตารางเข้าด้วยกันนั้น จะต้องมีการเรียงข้อมูลแต่ให้อยู่ในลำดับเดียวกันเสมอ เช่น A B C D ถ้าในฐานข้อมูลทรานแซ็คชันได้จัดเก็บเป็น B A จะต้องทาการจัดเรียงเป็น A B ก่อน ในทุกๆ ทรานแซ็คชัน โดยการทำนี้จะต้องเริ่มจากการหากลุ่มข้อมูลที่มีกลุ่ม 1 กลุ่มข้อมูลก่อน (k=1) เสมอ จึงเริ่มทำการค้นหา และในการรวมตารางนั้นจะต้องมีกลุ่มของข้อมูลส่วนหน้าเหมือนกันทั้งหมด ยกเว้นข้อมูลสุดท้ายในทุก ๆ k กลุ่มข้อมูล

ขั้นตอนที่สอง คือ การตัดตัวแทนหรือข้อมูลที่ไม่เป็นกลุ่มข้อมูลที่เกิดขึ้นบ่อยออก (Prune) เนื่องจากว่ากระบวนการรวมตารางเข้าด้วยกันนั้นทำให้เกิดกลุ่มของข้อมูลคู่แข่งขึ้นเป็นจำนวนมาก โดยหลักการทำนี้มีหลักการ คือ ถ้าซับเซตใด ๆ ของกลุ่มของข้อมูลคู่แข่งไม่อยู่ใน Lk −1 จะไม่ถูกนำมาพิจารณาให้ไปอยู่ใน Ck

โดยกระบวนการทำของออไพออรีในตัวอย่างการทำดังรูปที่ 2.11 โดยรูปนี้ได้ใช้ตารางจากข้อมูลของทรานแซ็คชันในรูปที่ 2.10



รูปที่ 2.11 กระบวนการหารูปแบบที่เกิดขึ้นบ่อยโดยใช้แนวคิดของออไพออรี

โดยหลังจากไม่สามารถรวมตารางของตัวแทนได้แล้วเราจะนำกลุ่มข้อมูลที่เกิดขึ้นบ่อยในทุก ๆ Lk ซึ่งถ้าจากตัวอย่างในรูปที่ 2.11 เราจะใช้กลุ่มข้อมูลทุกตัวในทุกๆ Lk ซึ่งในที่นี้ตั้งแต่ k=1 ถึง k=3 ซึ่งหลังจากได้กลุ่มข้อมูลที่เกิดขึ้นบ่อยเหล่านี้เราจะนำมาสร้างกฎของการเชื่อมโยงกัน โดยถ้ามี 2 กลุ่มข้อมูล จะสร้างกฎของการเชื่อมโยงได้ 2 กฎ และถ้ามี 3 กลุ่มข้อมูลจะสร้างได้ 6 กฎ ถ้า k กลุ่มข้อมูลโดยที่ k มากกว่าและเท่ากับ 2 จะสร้างกฎได้สองยกกำลังกลุ่มข้อมูล และนำมาลบด้วย 2

หลังจากที่ทำการเลือกกลุ่มข้อมูลที่เกิดขึ้นบ่อยแล้วทาการสร้างกฎการเชื่อมโยงของกลุ่มข้อมูลแต่ละตัวแล้ว จะต้องมีการคัดเลือกกฎของการเชื่อมโยงที่โดดเด่นโดยใช้ค่าความเชื่อมั่นตามรูปที่ 2.12 ซึ่งคือความเป็นไปได้ที่ทรานแซ็คชันที่เกิดข้อมูลอย่างหนึ่งขึ้นแล้วจะเกิด ซึ่งหาจากกลุ่มข้อมูลที่เกิดขึ้นพร้อมกันหารด้วยข้อมูลที่เกิดขึ้นอย่างเดียว เช่น A -> B ถ้าจะหาค่าความเชื่อมั่นก็หาจากจำนวนกลุ่มข้อมูลที่ A เกิดขึ้นพร้อมกับ B หารด้วย จำนวนของกลุ่มข้อมูลที่เกิด A อย่างเดียว โดยการที่เราจะพิจารณาว่ากฎการเชื่อมโยงใดเป็นกฎที่โดดเด่นจะต้องมีค่าความเชื่อมั่นมากกว่าค่าความเชื่อมั่นที่ต่ำที่สุดที่ยอมรับได้ (minimum confidence threshold) ถ้าหากกฎการเชื่อมโยงใดมีค่าความเชื่อมั่นมากกว่าค่าที่ยอมรับได้ก็จะรับการพิจารณาเป็นกฎการเชื่อมโยงที่โดดเด่น

$$confidence(A \Rightarrow B) = P(B|A) = \frac{support\_count(A \cup B)}{support\_count(A)}.$$

รูปที่ 2.12 สมการหาค่าความมั่นใจ

แม้ว่าเมื่อสามารถค้นหากลุ่มของข้อมูลที่เกิดขึ้นบ่อยครั้งจากข้อมูลที่มีอยู่มาได้และสามารถบ่งบอกกฎระหว่างกลุ่มออกมาได้ อันเป็นการแสดงว่ากลุ่มของข้อมูลเหล่านั้นอาจจะมีความสัมพันธ์กันอยู่ไม่มากก็น้อย ซึ่งในแต่ละกลุ่มนั้นเมื่อทำการสร้างกฎเกณฑ์ต่าง ๆ ระหว่างกลุ่มออกมาแล้วว่าแต่ละกลุ่มนั้นมีกฎเกณฑ์ต่อกันที่เข้มงวดมากขนาดไหน ยกตัวอย่างเช่น เมื่อมีกลุ่มแรกเกิดขึ้นเช่นนี้จะทำให้เกิดกลุ่มที่สองตามมาเสมอเช่นนี้ไป แต่อาจจะมีโอกาสที่ก่อให้เกิดกลุ่มที่สาม แต่ในความเป็นจริงแล้วข้อมูลกลุ่มแรกอาจมีความสัมพันธ์ต่อกลุ่มที่สองน้อยกว่ามีความสัมพันธ์ต่อกลุ่มที่สาม ซึ่งสิ่งเหล่านี้แสดงให้เห็นว่ากฎที่เข้มงวดต่อกันอาจจะไม่มีความสัมพันธ์กันมากก็เป็นไปได้ ซึ่งจำเป็นต้องทำการค้นหาต่อไปว่าข้อมูลใดที่มีความสัมพันธ์กันจริง และ ถึงแม้ว่าข้อมูลเหล่านั้นจะมีความสัมพันธ์กันก็ต้องค้นหาต่อไปว่าข้อมูลที่เราพบว่ามีความสัมพันธ์กันนั้นมีความสัมพันธ์กันในด้านบวกหรือว่าด้านลบ

ถึงแม้ว่าจะได้กฎต่าง ๆ มากมายจากการวิเคราะห์ข้อมูลทั้งหมด แต่ก็เป็นไปได้ว่ากฎทุกกฎที่ได้ค้นพบมานั้นใช่ว่าจะสำคัญเสมอไป สังเกตได้จากหัวข้อที่กล่าวผ่านมาแล้วนั้นถึงแม้ว่าแต่ละกลุ่มที่

ได้มานั้นจะมีกฎที่เข้มงวดต่อกันแต่ก็แสดงให้เห็นว่าอาจไม่มีความสัมพันธ์กันและไม่มีประโยชน์ที่จะนำกฎดังกล่าวที่ค้นพบเพื่อนำเอามาใช้ต่อไปในงาน ซึ่งเป็นหน้าที่ขั้นต่อไปที่จะต้องทำการคัดสรรต่อไปว่ากฎไหนหรือกฎข้อใดจากกลุ่มข้อมูลกลุ่มไหนที่ควรนำเอามาใช้ประโยชน์ต่อไป

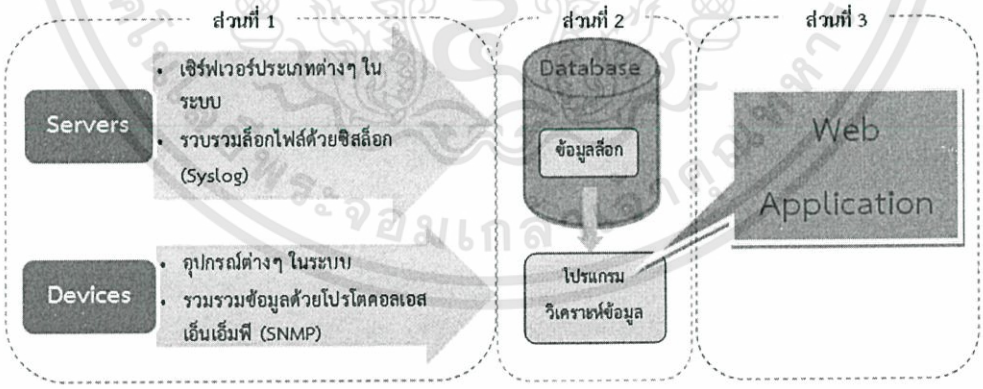
บทที่ 3
การออกแบบและพัฒนา

เนื่องจากโครงงานโปรแกรมตรวจพินิจการใช้ทรัพยากรระบบสารสนเทศอัตโนมัติเป็นโครงงานที่มีขนาดใหญ่ จำเป็นต้องมีการวางแผนและออกแบบโครงสร้างของระบบทั้งหมดก่อนพัฒนาในแต่ละส่วน ดังนี้

3.1 ภาพรวมของระบบ

โครงงานนี้ได้มีการออกแบบและพัฒนาระบบหลัก ๆ เป็นสามส่วน คือ

1) ส่วนแรก : ส่วนการรวบรวมข้อมูล เป็นการเก็บรวบรวมข้อมูลการทำงานของระบบและรายละเอียดการใช้งานทรัพยากรของระบบโดยมีการจัดเก็บลงฐานข้อมูลเดียวกัน
2) ส่วนที่สอง : ส่วนการวิเคราะห์ข้อมูล เป็นการนำข้อมูลจากฐานข้อมูลมาวิเคราะห์ด้วยโปรแกรมที่ทางกลุ่มวิจัยได้พัฒนาขึ้น
3) ส่วนที่สาม : ส่วนการแสดงผล เป็นการนำผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลมานำเสนอเพื่อให้ผู้ดูแลระบบสามารถนำไปตรวจสอบความผิดปกติที่เกิดขึ้นในระบบ โดยนำเสนอในรูปแบบของเว็บแอพพลิเคชัน



รูปที่ 3.1 ภาพรวมระบบ

3.1.1 รายละเอียดของ Environment ที่ใช้ในระบบ

ระบบโดยรวมจะมีบ้างส่วนที่จำเป็นต้องทำงานเฉพาะโปรแกรมซึ่งได้รายละเอียดของทรัพยากรที่ใช้มีดังต่อไปนี้

1) โปรแกรม Virtual box : เพื่อใช้ในการจำลองการทำงานของระบบเครือข่าย
2) Google Chrome : เพื่อใช้ในการแสดงผลลัพธ์ของข้อมูล

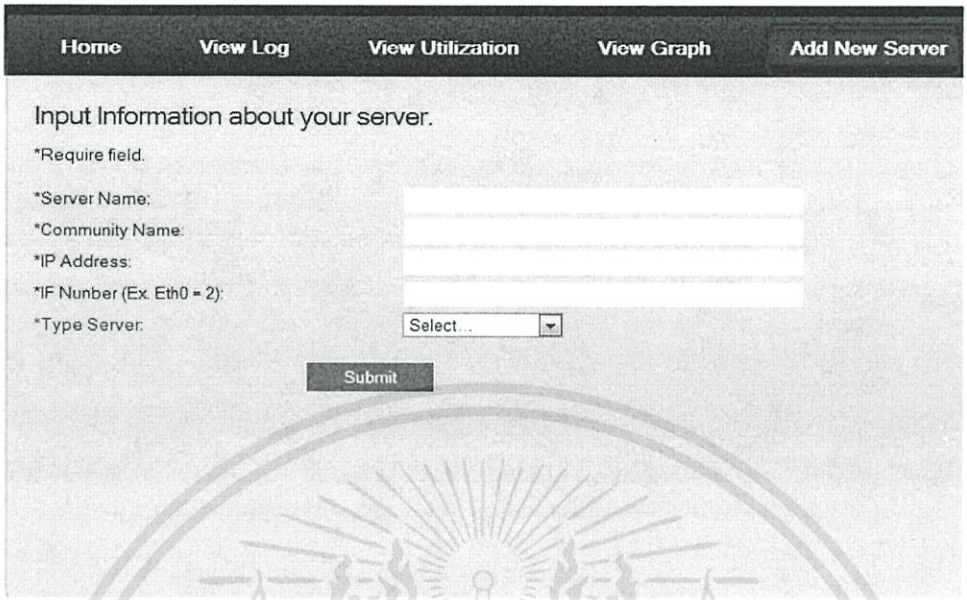
3.2 การออกแบบระบบ

ในส่วนของการออกแบบส่วนของระบบจะเป็นการแสดงผลในรูปแบบของเว็บแอพพลิเคชันซึ่งในหน้าเว็บนั้นจะมีการเพิ่มเซิร์ฟเวอร์, แสดงผลรายละเอียดการทำงานของระบบ, แสดงผลการใช้ทรัพยากรระบบ และผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูล ดังต่อไปนี้

1) ส่วนการเพิ่มเซิร์ฟเวอร์

ในระบบเครือข่ายส่วนใหญ่เมื่อทำการขึ้นระบบเรียบร้อยพอใช้งานได้สักระยะหนึ่งอาจมีการขยายระบบ จึงจำเป็นต้องมีการเพิ่มเครื่องเซิร์ฟเวอร์ต่าง ๆ เข้าไปในระบบ ทำให้ต้องมีการออกแบบหน้าเว็บนี้เพื่อตอบสนองการขยายขนาดของระบบ โดยในแต่ละช่องที่ต้องกรอกข้อมูลคือ

- Server Name : กำหนดชื่อประจำตัวของแต่ละเซิร์ฟเวอร์ โดยห้ามซ้ำกัน เนื่องจากเป็นตัวอ้างอิงการประมวลผลและแสดงผล
- Community Name : เปรียบเสมือนเป็นรหัสผ่านในการเข้าใช้โปรโตคอลเอสเอ็นเอ็มพี
- IP Address : เลขที่อยู่ไอพีของเครื่องนั้นๆ
- IF Numeber : ลำดับของเลขอินเตอร์เฟส (Interface Number) เพื่อใช้สำหรับการดึงค่าแบนด์วิดท์
- Type Server : เลือกชนิดของเซิร์ฟเวอร์ซึ่งจะแบ่งเป็นเว็บเซิร์ฟเวอร์และดาต้าเบสเซิร์ฟเวอร์ เนื่องจากกระบวนการเก็บข้อมูลและการทำเหมืองข้อมูลของเซิร์ฟเวอร์แต่ละประเภทไม่เหมือนกัน



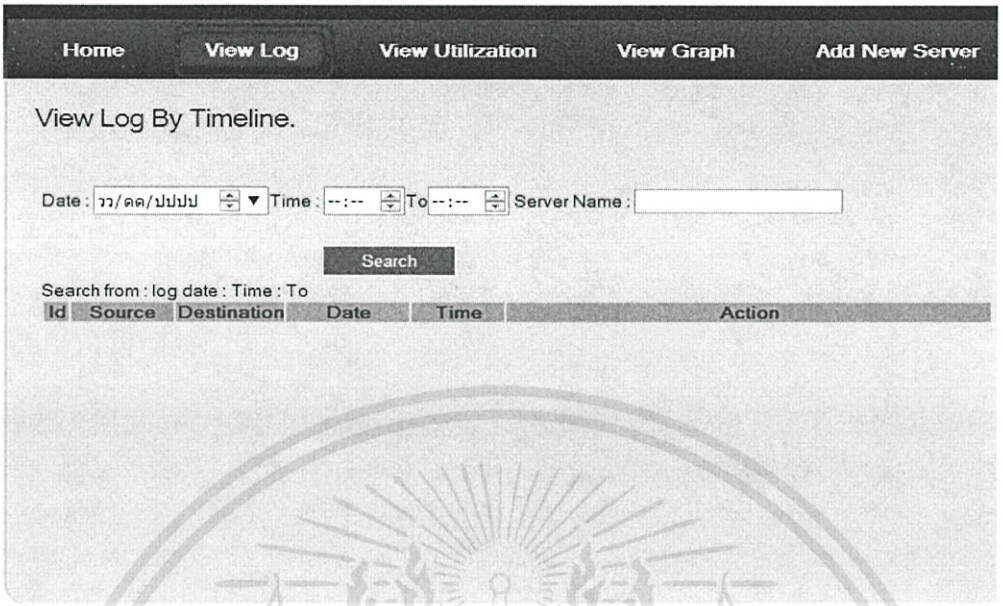
รูปที่ 3.2 การออกแบบหน้าเพิ่มเซิร์ฟเวอร์

2) ส่วนแสดงผลรายละเอียดการทำงานของระบบ

เป็นส่วนแสดงรายละเอียดเกี่ยวกับรายละเอียดของการทำงานภายในระบบจากเครื่องเซิร์ฟเวอร์แต่ละเครื่อง โดยมีการตั้งค่าต่าง ๆ ตามหัวข้อ ดังนี้

- Date : ตั้งค่าวันที่ต้องการ
- Time To : ตั้งค่าช่วงเวลาที่ต้องการ
- Server Name : ชื่อเซิร์ฟเวอร์ที่ต้องการดูผลการทำงาน

เมื่อกดปุ่ม Search ผลลัพธ์รายละเอียดการทำงานของระบบจะถูกนำมาแสดงผลในรูปแบบของตาราง



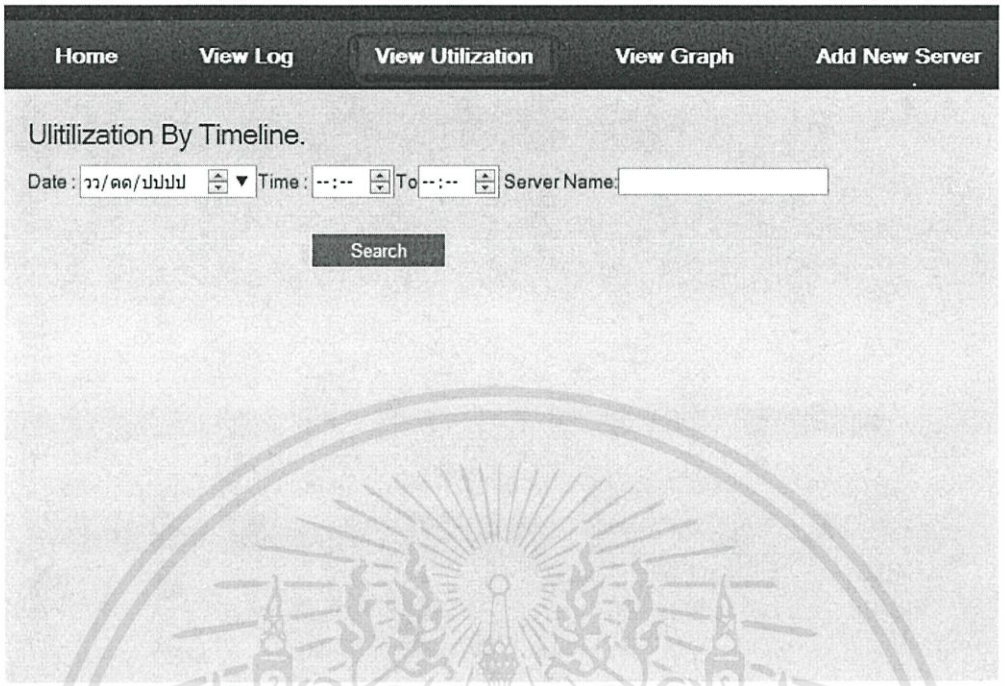
รูปที่ 3.3 การออกแบบหน้าแสดงรายละเอียดการทำงานของระบบ

3) ส่วนแสดงผลการใช้ทรัพยากรระบบ

เป็นส่วนแสดงถึงรายละเอียดการใช้ทรัพยากรของระบบเช่น ค่าการใช้ประโยชน์หน่วยประมวลผล (CPU Utilization), ค่าการใช้ประโยชน์หน่วยความจำ (Memory Utilization) เป็นต้น โดยมีการตั้งค่าต่าง ๆ ตามหัวข้อ ดังนี้

- Date : ตั้งค่าวันที่ต้องการ
- Time To : ตั้งค่าช่วงเวลาที่ต้องการ
- Server Name : ชื่อเซิร์ฟเวอร์ที่ต้องการดูผลการทำงาน

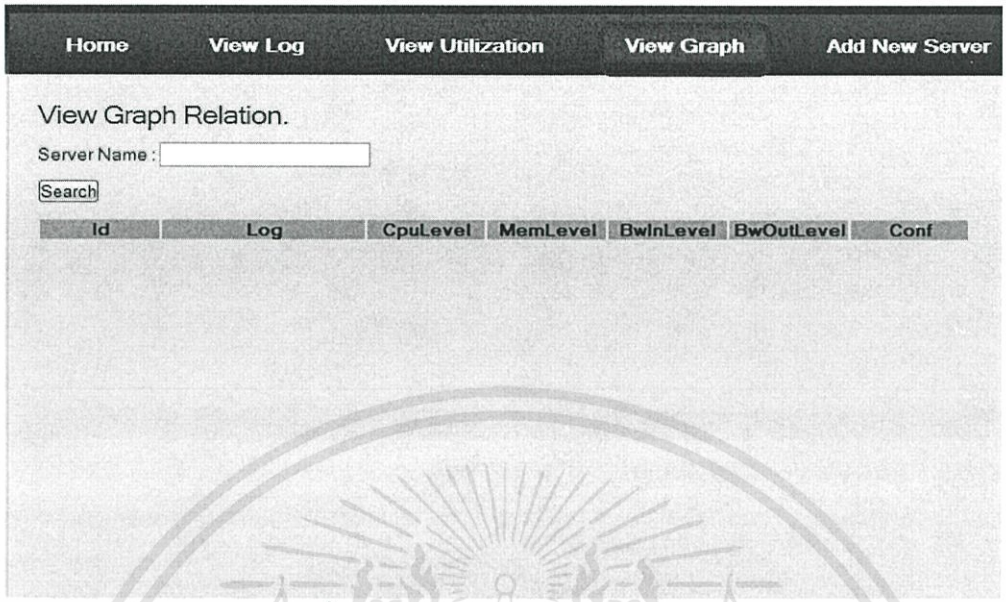
เมื่อกดปุ่ม Search ผลลัพธ์รายละเอียดการใช้งานทรัพยากรระบบทั้งหมดจะแสดงผลในรูปแบบของกราฟความสัมพันธ์



รูปที่ 3.4 การออกแบบหน้าแสดงผลการใช้ทรัพยากรระบบ

4) ส่วนแสดงผลลัพธ์ที่ได้จากการวิเคราะห์

เมื่อโปรแกรมได้ทำการวิเคราะห์ข้อมูลเรียบร้อยแล้ว นำมาแสดงผลในรูปของกราฟแสดงความสัมพันธ์ทางหน้าเว็บแอพลิเคชัน โดยสามารถกรอกชื่อเซิร์ฟเวอร์ที่ต้องการดูผลลัพธ์ซึ่งคาดว่ามีความผิดปกติเกิดขึ้นแล้วกดปุ่ม Search ผลลัพธ์จะปรากฏทั้งข้อมูลที่เป็นตารางผลลัพธ์ และแสดงเป็นกราฟความสัมพันธ์



รูปที่ 3.5 การออกแบบหน้าแสดงผลลัพธ์

3.3 การออกแบบฐานข้อมูล

เนื่องจากระบบการทำงานในเครือข่ายเป็นระบบที่มีขนาดใหญ่ จึงทำให้ต้องมีการวางแผนและออกแบบฐานข้อมูลเพื่อใช้ในการจัดเก็บข้อมูลที่ทำการรวบรวมจากเครื่องเซิร์ฟเวอร์แต่ละเครื่อง รวมถึงข้อมูลต่าง ๆ ที่จะนำไปใช้ในการวิเคราะห์หาความสัมพันธ์ โดยในส่วนของการเก็บรวบรวมข้อมูลในส่วนการใช้งานทรัพยากรระบบนั้น ได้แก่ ข้อมูลการใช้งานหน่วยประมวลผล, ข้อมูลการใช้งานหน่วยความจำ และข้อมูลการใช้งานแบนด์วิดท์ทั้งเข้าและออกนั้น จำเป็นต้องมีการการเปลี่ยนแปลงข้อมูลระดับการใช้งานดังกล่าวให้อยู่ในรูปแบบที่เหมาะสมสำหรับนำไปวิเคราะห์ โดยการเปลี่ยนระดับการใช้งานดังกล่าวให้อยู่ในรูปแบบของระดับ (Level) โดยจัดกลุ่มให้มีค่า 0 – 5 เนื่องจากระดับการใช้งานทรัพยากรมีค่าเป็นเปอร์เซ็น (Percent) ตั้งแต่ 0 – 100 ตามช่วงขณะเวลาต่าง ๆ เมื่อค่าระดับการใช้งานมีหลายค่าทำให้การวิเคราะห์ทำได้ยาก ซึ่งตารางต่าง ๆ ที่ใช้ในการเก็บข้อมูลมีการออกแบบฐานข้อมูลดังต่อไปนี้

3.3.1 การออกแบบตารางการเก็บข้อมูลการทำงานจากเว็บเซิร์ฟเวอร์

ข้อมูลการทำงานจากเว็บเซิร์ฟเวอร์ที่ทำการเก็บรวบรวมนั้นถือเป็นข้อมูลการทำงานของระบบอย่างหนึ่ง ดังนั้นข้อมูลที่จำเป็นจะต้องทำการรวบรวม มีดังต่อไปนี้

1) แหล่งที่มาของข้อมูล
2) วันที่และเวลาที่เกิดการทำงานของแต่ละเหตุการณ์ขึ้น

3) รายละเอียดของการทำงานที่เกิดขึ้น

ระบบจะทำการเก็บข้อมูลลงในตารางของฐานข้อมูลโดยตารางที่ทำหน้าที่เก็บข้อมูลรายละเอียดการทำงานสามารถอธิบายได้ดังรูปที่ 3.6

WEB LOGS	
PK	ID
	source
	destination
	date
	time
	action

รูปที่ 3.6 โครงสร้างตารางการเก็บข้อมูลการทำงานจากเว็บเซิร์ฟเวอร์

3.3.2 การออกแบบตารางการเก็บข้อมูลการทำงานจากดาต้าเบสเซิร์ฟเวอร์

ข้อมูลที่ทำการเก็บจากดาต้าเบสเซิร์ฟเวอร์ที่ทำการเก็บรวบรวมนั้นถือเป็นข้อมูลการทำงานของระบบอย่างหนึ่ง ดังนั้นข้อมูลที่จำเป็นจะต้องทำการรวบรวม มีดังต่อไปนี้

1) วันที่และเวลาที่เกิดการทำงานของแต่ละเหตุการณ์ขึ้น

2) รายละเอียดของการทำงานที่เกิดขึ้น

ซึ่งข้อมูลที่จำเป็นต้องเก็บรวบรวมลงในตารางของฐานข้อมูล โดยตารางที่ทำหน้าที่เก็บข้อมูลรายละเอียดดังกล่าวอธิบายได้ดังรูปที่ 3.7

DB LOGS	
PK	ID
	date
	time
	action

รูปที่ 3.7 โครงสร้างตารางการเก็บข้อมูลการทำงานจากดาต้าเบสเซิร์ฟเวอร์

3.3.3 การออกแบบตารางการเก็บข้อมูลการใช้งานทรัพยากรระบบ

การใช้งานทรัพยากรระบบในที่นี้หมายถึง ข้อมูลการใช้งานหน่วยประมวลผล, ข้อมูลการใช้งานหน่วยความจำ และข้อมูลการใช้งานแบนด์วิดท์ทั้งเข้าและออก ซึ่งจำเป็นต้องมีการการเปลี่ยนแปลงข้อมูลระดับการใช้งานดังกล่าวก่อนให้อยู่ในรูปแบบที่เหมาะสมสำหรับนำไปวิเคราะห์ โดยการเปลี่ยนระดับการใช้งานดังกล่าวให้อยู่ในรูปแบบของระดับ (Level) โดยจัดกลุ่มให้มีค่า 0 – 5 เนื่องจากระดับการใช้งานทรัพยากรมีค่าเป็นเปอร์เซ็นตั้งแต่ 0 – 100 ตามช่วงขณะเวลาต่าง ๆ เมื่อค่าระดับการใช้งานมีหลายค่าทำให้การวิเคราะห์ทำได้ยาก โดยระบบจะทำการจัดเก็บค่าต่าง ๆ ลงในตารางฐานข้อมูล ดังต่อไปนี้

1) แหล่งที่มาของข้อมูล (ชื่ออ้างอิงของอุปกรณ์ที่ถูกตรวจสอบระดับการใช้งานทรัพยากร)
2) วันที่และเวลาที่ทำการตรวจสอบระดับการใช้งานทรัพยากร
3) ระดับการใช้งานทรัพยากร

ระบบจะทำการเก็บข้อมูลลงในตารางของฐานข้อมูล โดยตารางที่ทำหน้าที่เก็บข้อมูลการใช้งานทรัพยากรระบบสามารถอธิบายได้ดังรูปที่ 3.8 – 3.11

CPU	
PK	ID
	source
	date
	time
	cpu
	cpu_level

รูปที่ 3.8 โครงสร้างตารางการเก็บรายละเอียดการใช้งานหน่วยประมวลผล

MEMORY	
PK	ID
	source
	date
	time
	mem
	mem_level

รูปที่ 3.9 โครงสร้างตารางการเก็บรายละเอียดการใช้งานหน่วยความจำ

BANDWIDTH_IN	
PK	ID
	source
	date
	time
	bandwidth
	bw_level

รูปที่ 3.10 โครงสร้างตารางการเก็บรายละเอียดการใช้งานแบนด์วิดท์ขาเข้า

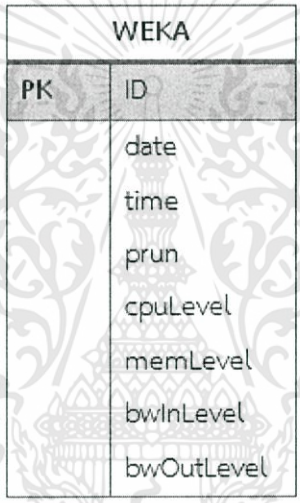
BANDWIDTH_OUT	
PK	ID
	source
	date
	time
	bandwidth
	bw_level

รูปที่ 3.11 โครงสร้างตารางการเก็บรายละเอียดการใช้งานแบนด์วิดท์ขาออก

3.3.4 การออกแบบตารางการเก็บข้อมูลที่ใช้ในการวิเคราะห์ความสัมพันธ์

เนื่องจากข้อมูลที่ทำการรวบรวมมาก่อนหน้านี้ คือ ข้อมูลการทำงานและการใช้งานทรัพยากรของระบบมารวมกัน เพื่อใช้ในการหาความสัมพันธ์ของข้อมูลดังกล่าว จึงจำเป็นต้องเก็บข้อมูลเฉพาะส่วนที่จำเป็นต่อการวิเคราะห์ความสัมพันธ์และจัดเก็บลงในตารางฐานข้อมูลอีกตารางหนึ่ง ซึ่งมีข้อมูลที่จำเป็นดังต่อไปนี้

1) วันที่และเวลาที่ทำการตรวจสอบระดับการทำงานและการใช้งานทรัพยากร
2) การกระทำของการทำงานของระบบที่ทำการกรองค่าแล้ว
3) ระดับการใช้งานทรัพยากรชนิดต่าง ๆ



รูปที่ 3.12 โครงสร้างตารางการเก็บข้อมูลที่ใช้ในการิเคราะห์ความสัมพันธ์

3.3.5 การออกแบบตารางการเก็บข้อมูลเพื่อใช้ในการแสดงผลลัพธ์

เนื่องจากผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลมานั้นต้องมีการจัดเก็บลงในตารางฐานข้อมูลเพื่อใช้ในการนำไปแสดงผลลัพธ์ต่อไป โดยระบบได้ทำการจัดเก็บค่าดังต่อไปนี้

1) การกระทำของการทำงานระบบ
2) ระดับการใช้งานทรัพยากรชนิดต่าง ๆ
3) ค่าความเชื่อมั่นของกฎความสัมพันธ์

VALUE	
PK	ID
	Log
	cpuLevel
	memLevel
	bwInLevel
	bwOutLevel
	Conf

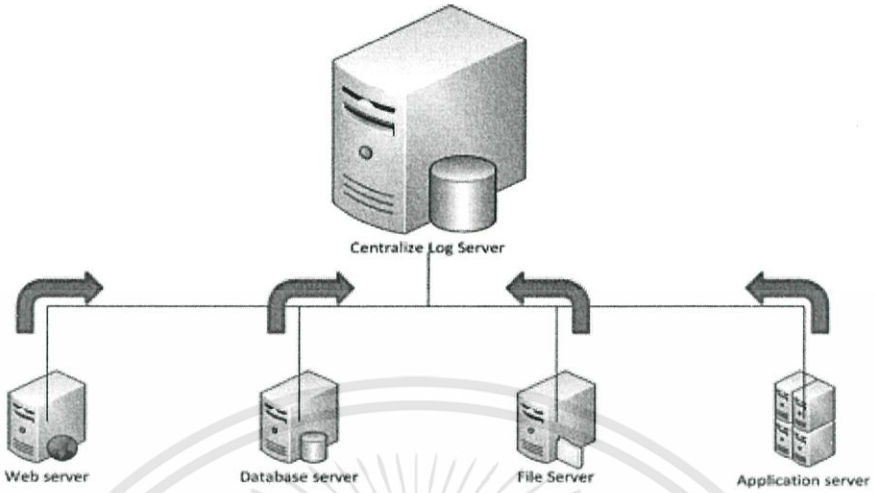
รูปที่ 3.13 โครงสร้างตารางการเก็บข้อมูลเพื่อใช้ในการแสดงผลลัพธ์

3.4 ส่วนการรวบรวมข้อมูล

ส่วนการรวบรวมข้อมูลนั้นได้แบ่งการจัดเก็บข้อมูลออกเป็นสองแบบด้วยกัน คือส่วนของการจัดเก็บรายละเอียดการทำงานของระบบ และส่วนของการจัดเก็บรายละเอียดการใช้งานทรัพยากรของระบบ

3.4.1 ส่วนการจัดเก็บรายละเอียดการทำงานของระบบ

ส่วนการจัดเก็บรายละเอียดข้อมูลการทำงานของระบบนั้น สามารถทำได้โดยการติดตั้งโปรแกรมซิสล็อก-เอ็นจี ซึ่งเป็นโปรแกรมที่สามารถรับและส่งข้อมูลรายละเอียดการทำงานต่าง ๆ ของระบบ อีกทั้งยังสามารถรวบรวมและจัดเก็บข้อมูลได้อย่างเป็นระบบ โดยโปรแกรมซิสล็อก-เอ็นจี จะดึงรายละเอียดการทำงานของระบบจากเซิร์ฟเวอร์แต่ละเครื่อง อย่างเช่น เว็บเซิร์ฟเวอร์ และดาต้าเบสเซิร์ฟเวอร์ แล้วนำไปจัดเก็บลงในฐานข้อมูลกลาง



รูปที่ 3.14 การรวบรวมข้อมูลการทำงานของระบบเพื่อเก็บลงฐานข้อมูล

3.4.2 ส่วนการจัดเก็บรายละเอียดการใช้งานทรัพยากรของระบบ

ส่วนของการจัดเก็บรายละเอียดการใช้งานทรัพยากรระบบ เช่น การใช้งานหน่วยประมวลผลของระบบและการใช้งานหน่วยความจำของระบบ เป็นต้นจะทำการรวบรวมโดยผ่านโปรโตคอลเอสเอ็นเอ็มพี ซึ่งเป็นโปรโตคอลที่ใช้ในการดึงข้อมูลการใช้งานทรัพยากรของเครื่องปลายทาง แล้วนำข้อมูลจากทั้งสองแหล่งจัดเก็บลงในฐานข้อมูลเดียวกัน ซึ่งการเก็บข้อมูลของทั้งสองส่วนสามารถอธิบายเป็นขั้นตอนได้ดังต่อไปนี้

ขั้นตอนที่หนึ่ง ทำการอ่านค่าระดับการใช้งานทรัพยากรในช่วงเวลาต่าง ๆ โดยช่วงเวลาที่ทำการอ่านค่าในระบบในแต่ละทรัพยากรจะใช้ช่วงเวลาที่แตกต่างกัน ซึ่งถ้าเป็นหน่วยประมวลผลและหน่วยความจำจะใช้ช่วงเวลาทุก ๆ 10 วินาที หน่วยความจำจะใช้ช่วงเวลาทุก ๆ 15 วินาที ในการอ่านค่าข้อมูล

ขั้นตอนที่สอง เมื่อทำการอ่านค่าระดับการใช้งานทรัพยากรของระบบเรียบร้อยแล้วจะทำการบันทึกเก็บลงไฟล์ เพื่อที่นำข้อมูลที่อ่านนั้นส่งไปเก็บยังฐานข้อมูลเดียวกับที่เก็บรายละเอียดการทำงานของระบบ

ขั้นตอนที่สาม เมื่อฐานข้อมูลของระบบได้รับข้อมูลรายละเอียดของระดับการใช้งานทรัพยากรของระบบก็จะทำการจัดเก็บลงฐานข้อมูลเพื่อนำไปใช้ในการวิเคราะห์ข้อมูลร่วมกับรายละเอียดการทำงานของระบบต่อไป

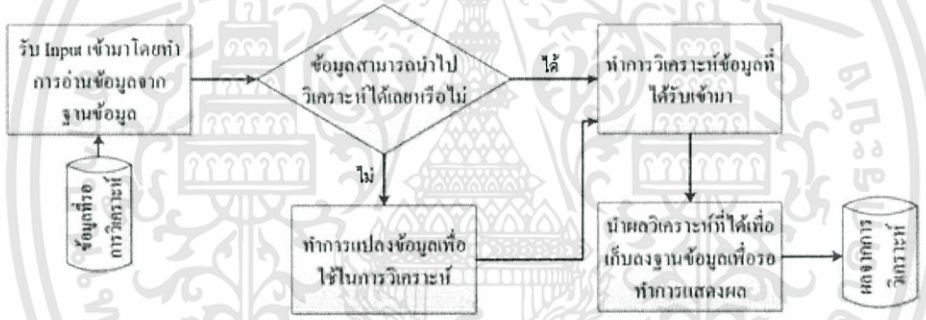
3.4.3 ส่วนการทำงานของฐานข้อมูล

ในส่วนของฐานข้อมูลจะทำการรับและเก็บรวบรวมรายละเอียดทั้งหมดภายในระบบ โดยรูปแบบการเก็บข้อมูลลงสู่ฐานข้อมูลจะทำการแยกออกเป็นสี่ส่วน ได้แก่

1) รายละเอียดการทำงานของระบบ
2) รายละเอียดการใช้งานทรัพยากรของระบบ
3) ตัวโปรแกรมที่ทำการวิเคราะห์ข้อมูล
4) ผลลัพธ์จากการวิเคราะห์ผ่านโปรแกรมตรวจวิเคราะห์

3.5 ส่วนการวิเคราะห์ข้อมูล

ส่วนของการนำข้อมูลที่ได้จากการรวบรวมมาจากการทำงานต่าง ๆ ในระบบ รวมถึงการใช้งานทรัพยากรของระบบ แล้วนำเข้าสู่โปรแกรมที่ทางกลุ่มผู้วิจัยได้พัฒนาขึ้นเพื่อทำการวิเคราะห์หาความสัมพันธ์กันของข้อมูล โดยโปรแกรมที่พัฒนาขึ้นนั้นได้พัฒนาด้วยภาษาจาวา ซึ่งมีการนำไลบารีของโปรแกรมวีก้า (Weka) ซึ่งเป็นโปรแกรมในการทำเหมืองข้อมูลมาใช้ในการสร้างโปรแกรมการวิเคราะห์หาความสัมพันธ์ของข้อมูล โดยใช้เทคนิคเรื่องกฎความสัมพันธ์ด้วยออไพรออรีอัลกอริทึม



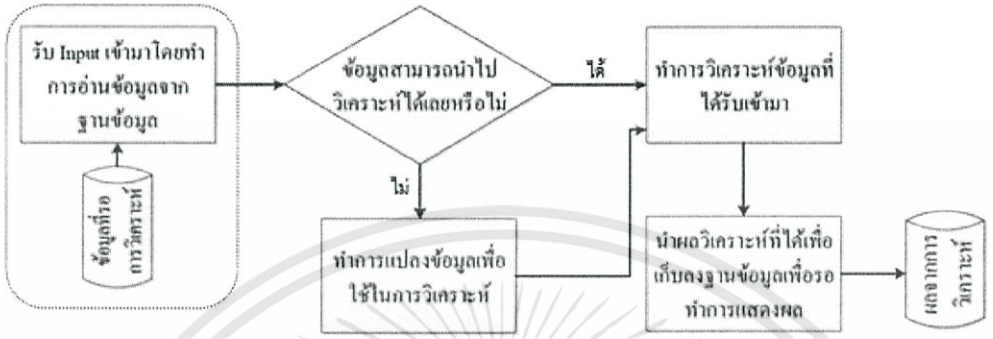
รูปที่ 3.15 ขั้นตอนการทำงานของโปรแกรมวิเคราะห์ข้อมูล

โปรแกรมที่ทำหน้าที่วิเคราะห์ข้อมูล แบ่งการทำงานของโปรแกรมออกเป็นสามส่วนประกอบไปด้วย ส่วนการนำเข้าข้อมูล ส่วนทำการวิเคราะห์ข้อมูล และส่วนจัดเก็บข้อมูลลงฐานข้อมูล

3.5.1 ส่วนการนำเข้าข้อมูล

โปรแกรมส่วนวิเคราะห์ข้อมูลจะทำการดึงข้อมูลจากฐานข้อมูลนำเข้าโปรแกรม ซึ่งข้อมูลทั้งหมดเป็นข้อมูลดิบที่ได้รับมาจากเครื่องในระบบโดยตรงแล้วเก็บลงฐานข้อมูล โดยข้อมูลที่นำเข้าสู่โปรแกรมจะถูกทำการตรวจสอบรูปแบบของข้อมูลที่นำเข้ามาก่อนว่าเป็นรูปแบบที่สามารถนำมาทำการวิเคราะห์ได้หรือไม่ ถ้าหากรูปแบบเกิดความผิดพลาดทำให้ยังไม่สามารถทำการวิเคราะห์ได้ทันที ตัวโปรแกรมจะต้องทำการปรับรูปแบบให้อยู่ในรูปแบบเดียวกันก่อน เพื่อให้ข้อมูลที่นำเข้ามามี

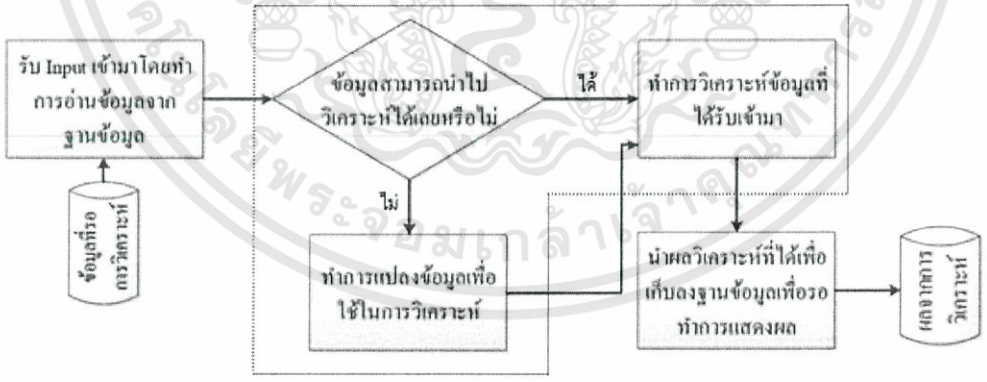
ลักษณะที่ตรงตามรูปแบบที่ตัวโปรแกรมต้องการเพื่อเข้าสู่กระบวนการวิเคราะห์หาความสัมพันธ์ของข้อมูล



รูปที่ 3.16 ส่วนการนำข้อมูลเข้าสู่โปรแกรม

3.5.2 ส่วนทำการวิเคราะห์ข้อมูล

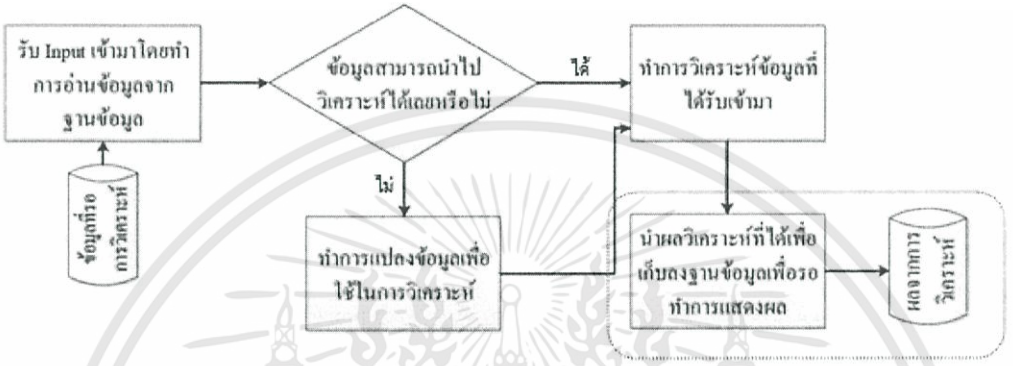
เมื่อโปรแกรมนำเข้าข้อมูลเรียบร้อยแล้ว โปรแกรมจะทำการวิเคราะห์ข้อมูล โดยทำการวิเคราะห์หาความสัมพันธ์ระหว่างการทำงานต่าง ๆ ของระบบที่มีผลกระทบต่อการใช้งานทรัพยากรของระบบ และทำการวิเคราะห์จำแนกประเภทของการทำงานเพื่อตรวจสอบว่าถ้ามีการทำงานในรูปแบบต่าง ๆ แล้วส่งผลต่อการใช้งานของทรัพยากรระบบอย่างไร



รูปที่ 3.17 ส่วนโปรแกรมทำการวิเคราะห์ข้อมูล

3.5.3 ส่วนการจัดเก็บข้อมูลลงฐานข้อมูล

เมื่อโปรแกรมทำการวิเคราะห์ข้อมูลเสร็จ จะทำการจัดเก็บผลลัพธ์ที่ได้นั้นลงในฐานข้อมูลเดิมที่ได้มีการดึงข้อมูลมาวิเคราะห์ โดยจัดเก็บแยกกันเป็นสัดส่วน เพื่อเตรียมข้อมูลนั้นไว้ในการนำไปแสดงผลต่อไป



รูปที่ 3.18 ส่วนจัดเก็บข้อมูลลงฐานข้อมูล

3.6 การตั้งค่าระบบ

การตั้งค่าระบบเพื่อทำการรวบรวมข้อมูลจะแบ่งออกเป็น 2 ส่วน คือ วิธีการติดตั้งซิสล็อก- เอ็นจีและวิธีการติดตั้งโปรโตคอลเอสเอ็นเอ็มพี และอีกส่วนหนึ่งคือเมื่อเริ่มต้นระบบต้องทำการตั้งค่าในส่วนของการวิเคราะห์ข้อมูล คือ จำเป็นต้องกรอกข้อมูลบางส่วนเพื่อทำการเริ่มต้นการวิเคราะห์ข้อมูล

- วิธีการติดตั้งและการใช้งานซิสล็อก – เอ็นจี เพื่อทำการรวบรวมข้อมูลการทำงานของระบบ แบ่งเป็นขั้นตอนต่าง ๆ ได้ดังนี้
  1) ติดตั้งซิสล็อก – เอ็นจี บนเครื่องเซิร์ฟเวอร์ที่ทำการจำลองไว้ในระบบเครือข่ายจำลองทุกเครื่อง
  2) เครื่องเซิร์ฟเวอร์ที่ไม่ใช่เครื่องเซิร์ฟเวอร์กลาง (Centralize Log) ให้ทำการตั้งค่าฟอเวิดล็อก (Forward Log) ของตนเองไปยังเครื่องเซิร์ฟเวอร์กลางผ่านทีซีพี พอร์ต (TCP Port) 2211 ไปยังเลขที่ไอพี (IP Address) 192.168.10.1 ซึ่งเป็นเลขที่ไอพีของเครื่องเซิร์ฟเวอร์กลาง

```
source s_apache {
        file("/var/log/apache2/access.log");
};

######
# destinations

destination d_net {
        tcp("192.168.10.1" port(2211));
};
log {
        source(s_apache);
        destination(d_net);
};
```

รูปที่ 3.19 การตั้งค่าฟอเวิดล็อกไปยังเซิร์ฟเวอร์กลาง

รายละเอียดของค่าแต่ละค่า มีดังนี้

Source คือ ต้นทางที่กำหนดทาง (Path) ของไฟล์ล็อกการทำงานของระบบ

Destination คือ ปลายทางที่ต้องการฟอเวิดข้อมูลออกไป

Log คือ การเชื่อมโยง (Map) ความสัมพันธ์ระหว่าง Source และ Destination

3) ที่เครื่องเซิร์ฟเวอร์กลางให้ตั้งค่าต้นทาง (Source) รับข้อมูลที่มาจากเซิร์ฟเวอร์อื่น ผ่านพอร์ต 2211 และค่าการใช้งานทรัพยากรที่ได้ทำการเก็บผ่านโปรโตคอล เอสเอ็นเอ็มพี

```
source web_cpu {
        file ("/etc/snmp/cpu-web.log" flags(no-parse) );
};
source web_mem {
        file ("/etc/snmp/memory-web.log" flags(no-parse) );
};
source bw_web_out {
        file ("/etc/snmp/bw-web-o.log" flags(no-parse) );
};
source bw_web_in {
        file ("/etc/snmp/bw-web-i.log" flags(no-parse) );
};
source net_log {
        tcp ( port(2211) flags(no-parse) );
};
```

รูปที่ 3.20 การตั้งค่าต้นทางเพื่อรับข้อมูลการทำงานและการใช้งานทรัพยากรที่มาจากเซิร์ฟเวอร์

4) เนื่องจากข้อมูลล็อกที่รับมานั้นมีส่วนที่ไม่จำเป็นต้องใช้งานด้วย จึงจำเป็นต้องทำการกรองข้อมูลนั้นออกเพื่อที่จะสามารถนำไปใช้ในกระบวนการทำเหมืองข้อมูลได้โดย

ไม่มีปัญหา การกรองนั้นทำได้โดยตัดแบ่งล็อกออกเป็นส่วน ๆ จากคอลัมน์ (Column), การเว้นวรรค หรือสัญลักษณ์ต่าง ๆ เช่น “ ”, [ ] เป็นต้น

```
parser p_apache {
        csv-parser( columns (
                                "CLIENT", "NAME", "USERNAME",
                                "TIMESTAMP", "URL", "STATUS",
                                "OUT", "REFERER", "AGENT", "A", "B", "C", "D")

                    flags(escape-double-char,strip-whitespace)
                    delimiters("    ")
                    quote-pairs('""[]')
                        );
};
parser p_mysql {
        csv-parser( columns (
                                "C1", "C2", "C3",
                               "C4" , "C5" )

                    flags(escape-double-char,strip-whitespace)
                    delimiters("       ")
                        );
};
```

รูปที่ 3.21 การตัดแบ่งล็อกออกเป็นส่วน ๆ

5) กำหนดปลายทาง (Destination) ไปยังฐานข้อมูลที่สร้างไว้เพื่อนำไปใช้ทำเหมืองข้อมูลต่อไป โดยจะประกอบด้วยเลขที่ไอพีของดาต้าเบสเซิร์ฟเวอร์ที่ทำการเก็บข้อมูล ชื่อ (Username) รหัสผ่าน (Password) ชื่อดาต้าเบส (Database Name) ชื่อตาราง (Tabale Name) ชื่อแอททริบิวต์ (Attribute Name) และค่า Value โดยค่า Value นั้นสามารถกำหนดได้จาก วัน เดือน ปี, เวลา, ข้อความ (String) ที่ได้จากการแบ่งล็อกออกเป็นส่วน ๆ ด้วยการทำพาเซอร์ (Parser) และค่าต่าง ๆ ที่จัดเก็บลงในฐานข้อมูลจะประกอบด้วยค่าการใช้งานหน่วยประมวลผล, หน่วยความจำ, แบนด์วิดท์ขาเข้า – ออก, ล็อกการทำงานของระบบ และ ล็อกที่ผ่านการกรองแล้ว

```
destination web_mysql_mem {
        sql(    type(mysql)
                host("localhost") username("root") password("admin")
                database("syslog")
                table("apachemem")
                columns("source","date","time","mem")
                values('$HOST','$YEAR-$MONTH-$DAY','$HOUR:$MIN:$SEC','${MEM}') );
};

destination apache_log {
        sql(    type(mysql)
                host("localhost") username("root") password("admin")
                database("syslog")
                table("apachelog")
                columns("source","destination","date","time","action")
                values('${URL}','$HOST','$YEAR-$MONTH-$DAY','$HOUR:$MIN:$SEC','${AGENT} ${C} ${D}') );

};
```

รูปที่ 3.22 การจัดเก็บค่าต่าง ๆ ลงไปยังตารางในฐานข้อมูล

6) ทำการเชื่อมโยง (Map) ค่าต้นทาง, พาเซอร์ และปลายทาง ให้สัมพันธ์กันด้วยฟังก์ชันล็อก

```
log {
        source(web_mem);
        parser(p_mem);
        destination(web_mysql_mem);
};

log {
        source(net_log);
        filter(f_favicon);
        filter(f_web);
        filter(f_ip);
        parser(p_apache);
        destination(apache_log);
};
```

รูปที่ 3.23 การเชื่อมโยงค่าด้วยฟังก์ชันล็อก

- วิธีการติดตั้งและการใช้งานโปรโตคอลเอสเอ็นเอ็มพี เพื่อทำการรวบรวมข้อมูลการใช้ทรัพยากรของระบบ แบ่งเป็นขั้นตอนต่าง ๆ ได้ดังนี้
  1) ติดตั้งโปรโตคอลเอสเอ็นเอ็มพีบนเครื่องเซิร์ฟเวอร์ที่ทำการลองในระบบเครือข่ายจำลองทุกเครื่อง
  2) ตั้งค่าชื่อกลุ่ม (Community Name) เพื่อกำหนด Group Local Network เพื่อสามารถทำการดึงค่าการใช้งานทรัพยากรผ่านทางโปรโตคอลเอสเอ็นเอ็มพี

```
com2sec local localhost bombie
com2sec network_1 192.168.10.0/24 bombie

##############################################

group MyG v1 local
group MyG v1 network_1
group MyG v1 network_2

##############################################

view all-mibs included .1 80

##############################################

access MyG "" v1 noauth exact all-mibs none none
```

รูปที่ 3.24 การตั้งค่าเพื่อเก็บข้อมูลการใช้งานทรัพยากร

3) เขียนคำสั่ง (Script) สำหรับดึงค่าการใช้งานหน่วยประมวลผลและหน่วยความจำ โดยคำสั่งนั้นจะทำการดึงค่าการใช้งานทุก ๆ 10 วินาที ผ่านโปรโตคอลเอสเอ็นเอ็มพี จากค่า MIB memTotalReal.0, memAvailReal.0, ssCpuUser.0, ssCpuSystem.0 และเก็บค่าอยู่ในรูปแบบของเปอร์เซ็นซึ่งได้จากการคำนวณ จากนั้นจะทำการบันทึกลงไฟล์ cpu-(SERVER NAME).log และ memory-(SERVER NAME).log

```
while(true)

do

memtotal=`snmpget -v 1 -c bombie 192.168.10.250 memTotalReal.0 | awk '{print $4}'`
memavail=`snmpget -v 1 -c bombie 192.168.10.250 memAvailReal.0 | awk '{print $4}'`
cpuuser=`snmpget -v 1 -c bombie 192.168.10.250 ssCpuUser.0 | awk '{print $4}'`
cpusystem=`snmpget -v 1 -c bombie 192.168.10.250 ssCpuSystem.0 | awk '{print $4}'`

x=100

tempmem=`expr ${memtotal} - ${memavail}`
tempmem1=`expr ${tempmem} \* ${x}`
percentmem=`expr ${tempmem1} / ${memtotal}`
percentcpu=`expr ${cpuuser} + ${cpusystem}`

echo $percentmem >> /etc/snmp/memory-web.log
echo $percentcpu >> /etc/snmp/cpu-web.log

sleep 10s
done
```

รูปที่ 3.25 คำสั่งสำหรับเก็บข้อมูลการใช้งานหน่วยประมวลผลและหน่วยความจำ

4) เขียนคำสั่งเพื่อดึงค่าการใช้งานแบนด์วิดท์ โดยคำสั่งจะทำการดึงค่าการใช้งานแบนด์วิดท์ทั้งขาเข้าและขาออกทุก ๆ 15 วินาที ผ่านค่า MIB MIB ifInOctets.X, ifOutOctets.X ซึ่งค่า X คือ ลำดับของ Interface จะเริ่มต้นด้วยค่า 2 เพราะ ค่า 1

คือ Loopback Interface และเนื่องจากค่า Traffic บน Network Interface นั้น จะทำการอัพเดต ทุก ๆ 15 วินาที ซึ่งเป็นค่าผลรวมของปริมาณการรับส่งข้อมูลทั้งหมด ต้องทำการรับค่าสองครั้งแล้วนำมาลบกันหารด้วย 15 เพื่อให้ได้ค่าเฉลี่ยต่อวินาที จากนั้นทำการจัดเก็บข้อมูลในหน่วย Bytes/s ลงใน File bw-(SERVER NAME)-i.log สำหรับแบนด์วิดท์ขาเข้า และ bw-(SERVER NAME)-o.log สำหรับแบนด์วิดท์ขาออก

```
while(true)
do
ifin=`snmpget -v 1 -c bombie 192.168.10.250 ifInOctets.2 | awk '{print $4}'`
ifout=`snmpget -v 1 -c bombie 192.168.10.250 ifOutOctets.2 | awk '{print $4}'`
sleep 15s
ifin2=`snmpget -v 1 -c bombie 192.168.10.250 ifInOctets.2 | awk '{print $4}'`
ifout2=`snmpget -v 1 -c bombie 192.168.10.250 ifOutOctets.2 | awk '{print $4}'`
x=15
totalin=`expr $ifin2 - $ifin`
totalin2=`expr ${totalin} / ${x}`
totalout=`expr $ifout2 - $ifout`
totalout2=`expr ${totalout} / ${x}`
echo $totalin2 >> /etc/snmp/bw-web-i.log
echo $totalout2 >> /etc/snmp/bw-web-o.log
done
```

รูปที่ 3.26 คำสั่งสำหรับเก็บข้อมูลการใช้งานแบนด์วิดท์

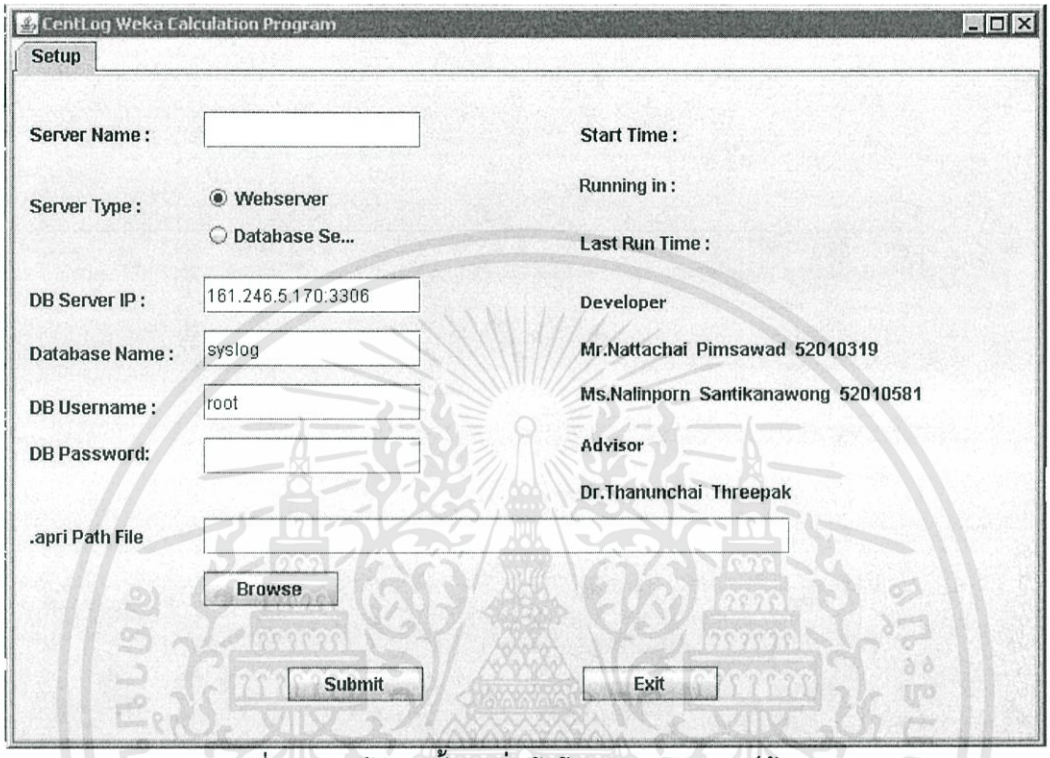
5) เขียนคำสั่งเพื่อสั่งให้คำสั่งนั้นดึงข้อมูลการใช้งานทรัพยากรทำงานแบบเป็นพื้นหลัง (Background) คือ การให้คำสั่งทำงานตลอดเวลาโดยอัตโนมัติ

```
#!/bin/bash

./cpu-mem-db.sh & > /dev/null
./cpu-mem-web.sh & > /dev/null
./bw-web.sh & > /dev/null
./bw-db.sh & > /dev/null
```

รูปที่ 3.27 คำสั่งเพื่อให้ระบบทำการเก็บข้อมูลการใช้งานตลอดเวลา

- วิธีตั้งค่าการเริ่มต้นการวิเคราะห์ข้อมูล เมื่อทำการรันเพื่อเก็บข้อมูลจากการส่วนการทำงานและการใช้งานทรัพยากรระบบ แล้วก็ต้องทำการรันในส่วนของโปรแกรมวิเคราะห์ข้อมูลตอนเริ่มต้นด้วย โดยจะมีหน้าต่างเพื่อทำการกรอกข้อมูล ชื่อเซิร์ฟเวอร์

ประเภทของเซิร์ฟเวอร์ หมายเลขที่อยู่ไอพีของดาต้าเบสที่เก็บข้อมูล ชื่อดาต้าเบส ชื่อผู้ใช้งาน และรหัสผ่านของดาต้าเบส



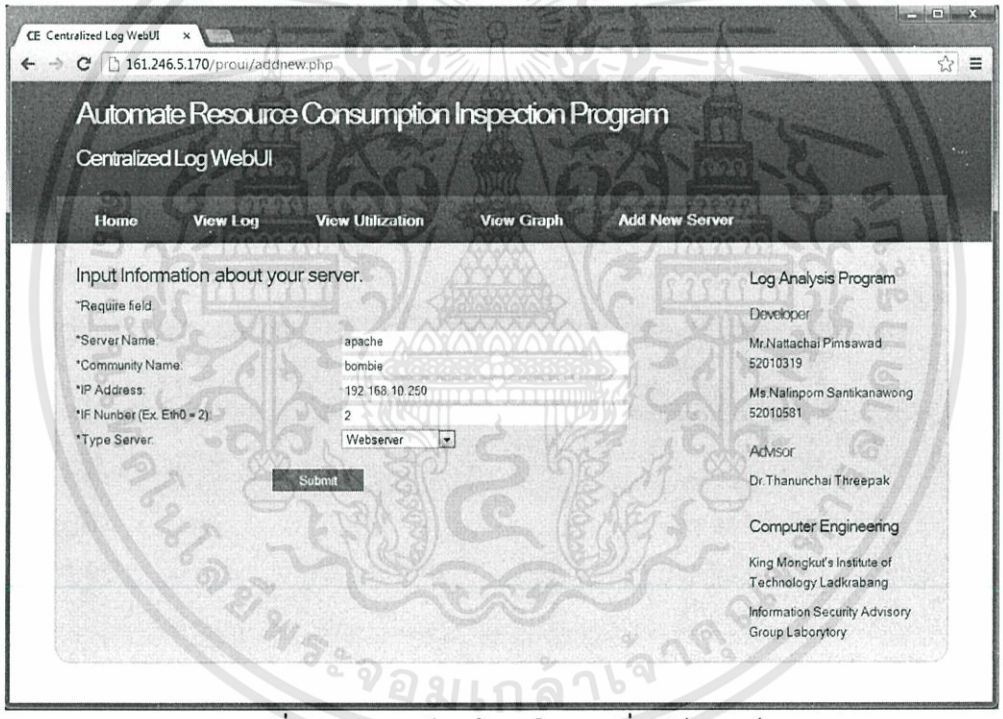
รูปที่ 3.28 หน้าการตั้งค่าเริ่มต้นโปรแกรมวิเคราะห์ข้อมูล

บทที่ 4
การทดลองและผลการทดลอง

4.1 การตั้งค่าระบบ

การเริ่มต้นระบบต้องมีการเพิ่มค่าเซิร์ฟเวอร์เข้าระบบก่อนที่จะทำการจัดเก็บ รวบรวมข้อมูลการทำงานของระบบได้ โดยจากหน้าเว็บแอพพลิเคชันจะมีหน้าหนึ่งที่ให้เพิ่มเซิร์ฟเวอร์ได้โดยมีขั้นตอนดังต่อไปนี้

1) กรอกข้อมูลรายละเอียดต่าง ๆ เกี่ยวกับเซิร์ฟเวอร์ที่ต้องการ ดังรูปที่ 4.1



รูปที่ 4.1 รายละเอียดข้อมูลในการเพิ่มเซิร์ฟเวอร์

2) เมื่อกดปุ่ม Submit แล้ว เว็บแอพพลิเคชันจะทำการสร้างตารางต่าง ๆ ที่จำเป็นต้องใช้เพื่อเก็บรวบรวมข้อมูล และสร้างทริกเกอร์ของมายเอสคิวแอล (Trigger My SQL) สำหรับกำหนดค่าระดับการใช้งานทรัพยากรระบบลงบนฐานข้อมูล

```
DROP TRIGGER IF EXISTS `apachecputrig`//
CREATE TRIGGER `apachecputrig` BEFORE INSERT ON `apachecpu`
 FOR EACH ROW BEGIN
IF NEW.CPU BETWEEN 0 AND 20 THEN SET NEW.cpu_level = 1 ;
ELSEIF NEW.CPU BETWEEN 21 AND 40 THEN SET NEW.cpu_level = 2 ;
ELSEIF NEW.CPU BETWEEN 41 AND 60 THEN SET NEW.cpu_level = 3 ;
ELSEIF NEW.CPU BETWEEN 61 AND 81 THEN SET NEW.cpu_level = 4 ;
ELSE SET NEW.cpu_level = 5 ;
END IF;
END
//
```

รูปที่ 4.2 การตั้งค่าทริกเกอร์มายเอสคิวแอลของการใช้งานหน่วยประมวลผล

การตั้งค่าทริกเกอร์มายเอสคิวแอลในแต่ละประเภทของทรัพยากรนั้น มีรูปแบบเหมือนกันโดยค่าที่มีการเปลี่ยนแปลงนั้น คือ ในส่วนของบรรทัดที่ 2 ในรูป และส่วนของค่าที่ต้องการแบ่งออกเป็นช่วง

Structure | SQL | Search | Query | Export | Import | Operations | Privileges | Drop

	Table	Action	Records [1]	Type	Collation	Size	Overhead
☐	apachebwin		0	MyISAM	latin1_swedish_ci	1.0 KiB	-
☐	apachebwout		0	MyISAM	latin1_swedish_ci	1.0 KiB	-
☐	apachecpu		0	MyISAM	latin1_swedish_ci	1.0 KiB	-
☐	apachelog		0	MyISAM	latin1_swedish_ci	1.0 KiB	-
☐	apachemem		0	MyISAM	latin1_swedish_ci	1.0 KiB	-
☐	apachevalue		0	MyISAM	latin1_swedish_ci	1.0 KiB	-
☐	apacheweka		0	MyISAM	latin1_swedish_ci	1.0 KiB	-
	7 table(s)	Sum	0	MyISAM	latin1_swedish_ci	7.0 KiB	0 B

รูปที่ 4.3 ตารางที่สร้างขึ้นในฐานข้อมูลเมื่อมีการเพิ่มเซิร์ฟเวอร์

ในส่วนของตารางที่สร้างเพื่อใช้เก็บรวบรวมข้อมูลหลังจากที่มีการเพิ่มเซิร์ฟเวอร์เข้าในระบบแล้วมีมีดังนี้

- Apachebwin : ตารางการเก็บค่าการใช้งานแบนด์วิดท์ขาเข้า
- Apachebwout : ตารางการเก็บค่าการใช้งานแบนด์วิดท์ขาออก
- Apachecpu : ตารางการเก็บค่าการใช้งานหน่วยประมวลผล
- Apachelog : ตารางการเก็บค่าการทำงานของเว็บเซิร์ฟเวอร์
- Apachemem : ตารางการเก็บค่าการใช้งานหน่วยความจำ
- Apachevalue : ตารางการเก็บค่าผลลัพธ์จากการวิเคราะห์ความสัมพันธ์
- Apacheweka : ตารางการเก็บค่าข้อมูลเพื่อใช้ในการนำไปวิเคราะห์หาความสัมพันธ์

3) หลังจากนั้นเว็บแอพพลิเคชันจะทำการสร้างไฟล์คำสั่ง (Script) มาจำนวน 4 ไฟล์ ได้แก่ apache-addayslog.sh, apache-bw.sh, apache-cpu-mem.sh และ run-apache.sh ดังรูป

```
-rw-r--r-- 1 www-data www-data    7740 2013-03-04 18:41 apache-addsyslog.sh
-rw-r--r-- 1 www-data www-data     582 2013-03-04 18:41 apache-bw.sh
-rw-r--r-- 1 www-data www-data     625 2013-03-04 18:41 apache-cpu-mem.sh
-rw-r--r-- 1 www-data www-data     136 2013-03-04 18:41 run-apache.sh
```

รูปที่ 4.4 รายละเอียดของไฟล์คำสั่งที่เว็บแอพพลิเคชันสร้างขึ้น

- Apche-addsyslog.sh เป็นไฟล์คำสั่งสำหรับการตั้งค่าในซิสล็อก-เอ็นจี

```
 GNU nano 2.2.2                File: apache-addsyslog.sh

#!/bin/bash
echo "source apache_cpu {" >> /opt/syslog-ng/etc/syslog-ng.conf
echo    "file ("'"'"/etc/snmp/apache-cpu.log"'"' "flags(no-parse) );" >> /opt/syslog-ng/etc/syslog-ng.conf
echo "};" >> /opt/syslog-ng/etc/syslog-ng.conf

echo "source apache_mem {" >> /opt/syslog-ng/etc/syslog-ng.conf
echo    "file ("'"'"/etc/snmp/apache-mem.log"'"' "flags(no-parse) );" >> /opt/syslog-ng/etc/syslog-ng.conf
echo "};" >> /opt/syslog-ng/etc/syslog-ng.conf

echo "source apache_bwin {" >> /opt/syslog-ng/etc/syslog-ng.conf
echo    "file ("'"'"/etc/snmp/apache-bw-i.log"'"' "flags(no-parse) );" >> /opt/syslog-ng/etc/syslog-ng.conf
echo "};" >> /opt/syslog-ng/etc/syslog-ng.conf

echo "source apache_bwout {" >> /opt/syslog-ng/etc/syslog-ng.conf
echo "file ("'"'"/etc/snmp/apache-bw-o.log"'"' "flags(no-parse) );" >> /opt/syslog-ng/etc/syslog-ng.conf
echo "};" >> /opt/syslog-ng/etc/syslog-ng.conf

echo "filter apache_ip {" >> /opt/syslog-ng/etc/syslog-ng.conf
echo    "host ("'"'"192.168.10.250"'"'");" >> /opt/syslog-ng/etc/syslog-ng.conf
echo "};" >> /opt/syslog-ng/etc/syslog-ng.conf

echo "destination apache_mysql_cpu {" >> /opt/syslog-ng/etc/syslog-ng.conf
echo    "sql("     "type(mysql)" >> /opt/syslog-ng/etc/syslog-ng.conf
echo          "host("'"'"localhost"'"'")" "username"("'"'"root"'"'")" "password"("'"'"admin"'"'")" >> /opt$
echo          "database"("'"'"syslog2"'"'")" >> /opt/syslog-ng/etc/syslog-ng.conf
echo          "table"("'"'"apachecpu"'"'")" >> /opt/syslog-ng/etc/syslog-ng.conf
echo          "columns"("'"'"source"'"'","'"'"date"'"'","'"'"time"'"'","'"'"cpu"'"'")" >> /opt/syslog-ng/et$
echo          "values("'"'"$""HOST"'"'","'"'"$""YEAR""-""$""MONTH""-""$""DAY"'"'","'"'"$""HOUR"":""$""MIN"":$
echo "}," >> /opt/syslog-ng/etc/syslog-ng.conf
```

รูปที่ 4.5 ไฟล์คำสั่ง Apache-addsyslog.sh

- Apache-bw.sh เป็นไฟล์คำสั่งสำหรับการดึงข้อมูลแบนด์วิดท์ผ่านทางโปรโตคอลเอสเอ็นเอ็มพี

```
GNU nano 2.2.2                    File: apache-bw.sh

#!/bin/bash
while(true)
do
ifin=`snmpget -v 1 -c bombie 192.168.10.250 ifInOctets.2 | awk '{print $4}'`
ifout=`snmpget -v 1 -c bombie 192.168.10.250 ifOutOctets.2 | awk '{print $4}'`
sleep 15s
ifin2=`snmpget -v 1 -c bombie 192.168.10.250 ifInOctets.2 | awk '{print $4}'`
ifout2=`snmpget -v 1 -c bombie 192.168.10.250 ifOutOctets.2 | awk '{print $4}'`
x=15
totalin=`expr $ifin2 - $ifin`
totalin2=`expr ${totalin} / ${x}`
totalout=`expr $ifout2 - $ifout`
totalout2=`expr ${totalout} / ${x}`
echo $totalin2 >> /etc/snmp/apache-bw-i.log
echo $totalout2 >> /etc/snmp/apache-bw-o.log
done
```

รูปที่ 4.6 ไฟล์คำสั่ง Apache-bw.sh

- Apache-cpu-mem.sh เป็นไฟล์คำสั่งสำหรับการดึงข้อมูลการใช้งานหน่วยประมวลผลและหน่วยความจำผ่านทางโปรโตคอลเอสเอ็นเอ็มพี

```
GNU nano 2.2.2                    File: apache-cpu-mem.sh

#!/bin/bash
while(true)
do
memtotal=`snmpget -v 1 -c bombie 192.168.10.250 memTotalReal.0 | awk '{print $4}'`
memavail=`snmpget -v 1 -c bombie 192.168.10.250 memAvailReal.0 | awk '{print $4}'`
cpuuser=`snmpget -v 1 -c bombie 192.168.10.250 ssCpuUser.0 | awk '{print $4}'`
cpusystem=`snmpget -v 1 -c bombie 192.168.10.250 ssCpuSystem.0 | awk '{print $4}'`
x=100
tempmem=`expr $memtotal - $memavail`
tempmem1=`expr ${tempmem} \* ${x}`
percentmem=`expr ${tempmem1} / ${memtotal}`
percentcpu=`expr ${cpuuser} + ${cpusystem}`
echo $percentmem >> /etc/snmp/apache-mem.log
echo $percentcpu >> /etc/snmp/apache-cpu.log
sleep 10s
done
```

รูปที่ 4.7 ไฟล์คำสั่ง Apache-cpu-mem.sh

- Run-apache.sh เป็นไฟล์คำสั่งสำหรับสั่งให้เริ่มการทำงานทั้งหมด คือ รีสตาร์ท (Restart) ซิสล็อก – เอ็นจี ที่ได้ถูกตั้งค่าเข้าไปใหม่ และเริ่มต้นเก็บค่าการใช้งานทรัพยากรระบบผ่านทางโปรโตคอลเอสเอ็นเอ็มพี

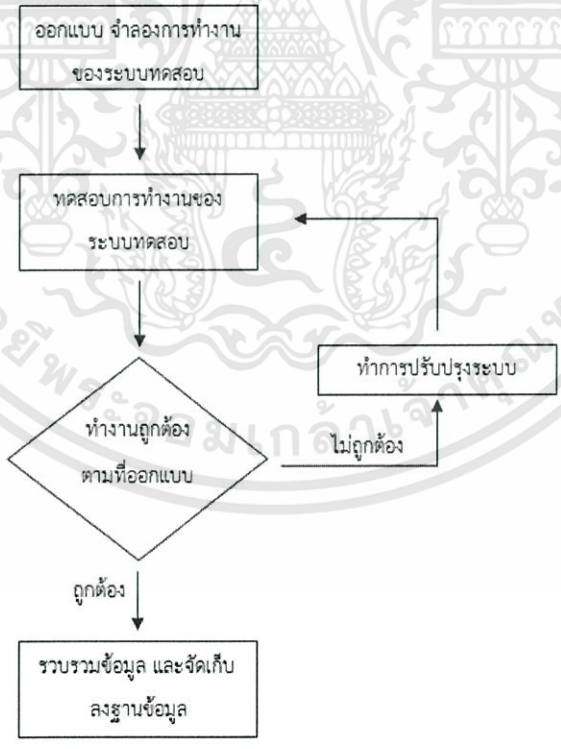
```
 GNU nano 2.2.2                          File: run-apache.sh

#!/bin/bash
bash apache-cpu-mem.sh & > /dev/null
bash apache-bw.sh & > /dev/null
bash apache-addsyslog.sh
/etc/init.d/syslog-ng restart
```

รูปที่ 4.8 ไฟล์คำสั่ง Run-apache.sh

4) เมื่อสั่งรัน (Run) คำสั่ง run-apache.sh ระบบต่าง ๆ จะเริ่มทำงานทั้งการเก็บค่าการทำงานและการใช้งานของระบบลงฐานข้อมูล ดังรูปที่ 4.10 – 4.15

4.2 การจัดเก็บรายละเอียดของระบบ

ในส่วนของการจัดเก็บรายละเอียดของระบบแบ่งออกเป็นสองส่วนหลัก ๆ คือ ส่วนของการจัดเก็บรายละเอียดการทำงานของระบบ และอีกส่วนเป็นการจัดเก็บรายละเอียดการใช้งานทรัพยากรของระบบ ซึ่งมีขั้นตอนของการจัดเก็บ ดังรูปที่ 4.9



รูปที่ 4.9 ขั้นตอนการจัดเก็บรายละเอียดของระบบ

ขั้นตอนการจัดเก็บรายละเอียดของระบบนั้นเริ่มจากกออกแบบระบบทดสอบก่อน แล้วทำการจำลองระบบที่ได้ออกแบบในโปรแกรมจำลองการทำงาน เช่น โปรแกรม Virtual Box และเมื่อได้ทำการจำลองระบบเรียบร้อยจึงทำการตรวจสอบว่าระบบสามารถทำงานได้ตามที่ออกแบบหรือไม่ ถ้าหากไม่ถูกต้องก็ทำการปรับปรุงระบบใหม่ ถ้าหากระบบที่จำลองสามารถทำงานได้ถูกต้องแล้วจึงทำการรวบรวมข้อมูลและจัดเก็บรายละเอียดการทำงานนั้นลงฐานข้อมูล เพื่อนำไปใช้การวิเคราะห์หาความสัมพันธ์ต่อไป

4.2.1 การจัดเก็บรายละเอียดการทำงานของระบบ

ในส่วนของการจัดเก็บรายละเอียดการทำงาน สามารถทำได้โดยการติดตั้งโปรแกรมซิสล็อก-เอ็นจี เพื่อใช้ในการรวบรวมล็อกไฟล์การทำงานของระบบและส่งไปเก็บยังตารางที่ได้ออกแบบไว้แล้วในฐานข้อมูล

id	source	destination	date	time	action
1	192.168.10.254	192.168.10.250	2013-03-02	03:32:41	GET/index.html - Mozilla/5.0 (Windows NT 5.1) Appl...
2	192.168.10.254	192.168.10.250	2013-03-02	03:32:42	GET/index2.html - Mozilla/5.0 (Windows NT 5.1) App...
3	192.168.10.254	192.168.10.250	2013-03-02	03:32:43	GET/test.php - Mozilla/5.0 (Windows NT 5.1) AppleW...
4	192.168.10.254	192.168.10.250	2013-03-02	03:32:46	GET/normaldb.php - Mozilla/5.0 (Windows NT 5.1) Ap...
5	192.168.10.254	192.168.10.250	2013-03-02	03:33:10	GET/dbpeak.php - Mozilla/5.0 (Windows NT 5.1) Appl...
6	192.168.10.254	192.168.10.250	2013-03-02	03:33:10	GET/webpeak.php - Mozilla/5.0 (Windows NT 5.1) App...
7	192.168.10.254	192.168.10.250	2013-03-02	03:34:24	GET/index.html - Mozilla/5.0 (Windows NT 5.1) Appl...
8	192.168.10.254	192.168.10.250	2013-03-02	03:34:25	GET/index2.html - Mozilla/5.0 (Windows NT 5.1) App...
9	192.168.10.254	192.168.10.250	2013-03-02	03:34:26	GET/test.php - Mozilla/5.0 (Windows NT 5.1) AppleW...
10	192.168.10.254	192.168.10.250	2013-03-02	03:34:27	GET/normaldb.php - Mozilla/5.0 (Windows NT 5.1) Ap...
11	192.168.10.254	192.168.10.250	2013-03-02	03:34:53	GET/webpeak.php - Mozilla/5.0 (Windows NT 5.1) App...
12	192.168.10.254	192.168.10.250	2013-03-02	03:35:00	GET/dbpeak.php - Mozilla/5.0 (Windows NT 5.1) Appl...
13	192.168.10.254	192.168.10.250	2013-03-02	03:35:29	GET/index.html - Mozilla/5.0 (Windows NT 5.1) Appl...
14	192.168.10.254	192.168.10.250	2013-03-02	03:35:30	GET/index2.html - Mozilla/5.0 (Windows NT 5.1) App...
15	192.168.10.254	192.168.10.250	2013-03-02	03:35:31	GET/test.php - Mozilla/5.0 (Windows NT 5.1) AppleW...
16	192.168.10.254	192.168.10.250	2013-03-02	03:35:32	GET/normaldb.php - Mozilla/5.0 (Windows NT 5.1) Ap...
17	192.168.10.254	192.168.10.250	2013-03-02	03:35:58	GET/webpeak.php - Mozilla/5.0 (Windows NT 5.1) App...
18	192.168.10.254	192.168.10.250	2013-03-02	03:36:05	GET/dbpeak.php - Mozilla/5.0 (Windows NT 5.1) Appl...
19	192.168.10.254	192.168.10.250	2013-03-02	03:38:40	GET/index.html - Mozilla/5.0 (Windows NT 5.1) Appl...
20	192.168.10.254	192.168.10.250	2013-03-02	03:38:49	GET/index2.html - Mozilla/5.0 (Windows NT 5.1) App...

รูปที่ 4.10 ตัวอย่างข้อมูลรายละเอียดการทำงานของเว็บเซิร์ฟเวอร์

id	date	time	action
1	2013-03-02	03:32:32	73 Query select student1.ID,student2.SURNAME,stude...
2	2013-03-02	03:32:46	74 Query SELECT * FROM student3
3	2013-03-02	03:34:22	75 Query select student1.ID,student2.SURNAME,stude...
4	2013-03-02	03:34:27	76 Query SELECT * FROM student3
5	2013-03-02	03:35:27	77 Query select student1.ID,student2.SURNAME,stude...
6	2013-03-02	03:35:33	78 Query SELECT * FROM student3
7	2013-03-02	03:38:14	79 Query select student1.ID,student2.SURNAME,stude...
8	2013-03-02	03:39:11	80 Query SELECT * FROM student3
9	2013-03-02	03:43:55	81 Query select student1.ID,student2.SURNAME,stude...
10	2013-03-02	03:48:23	82 Query SELECT * FROM student3
11	2013-03-02	03:49:34	83 Query select student1.ID,student2.SURNAME,stude...
12	2013-03-02	03:55:14	84 Query select student1.ID,student2.SURNAME,stude...
13	2013-03-02	03:57:33	85 Query SELECT * FROM student3
14	2013-03-02	04:00:51	86 Query select student1.ID,student2.SURNAME,stude...
15	2013-03-02	04:06:28	87 Query select student1.ID,student2.SURNAME,stude...
16	2013-03-02	04:06:44	88 Query SELECT * FROM student3
17	2013-03-02	04:12:04	89 Query select student1.ID,student2.SURNAME,stude...
18	2013-03-02	04:15:55	90 Query SELECT * FROM student3
19	2013-03-02	04:17:42	91 Query select student1.ID,student2.SURNAME,stude...
20	2013-03-02	04:23:20	92 Query select student1.ID,student2.SURNAME,stude...

รูปที่ 4.11 ตัวอย่างข้อมูลรายละเอียดการทำงาของดาต้าเบสเซิร์ฟเวอร์

4.2.2 การจัดเก็บรายละเอียดการใช้งานทรัพยากรของระบบ

ส่วนของการรวบรวมการใช้งานทรัพยากรของระบบสามารถทำได้โดยใช้โปรโตคอลเอสเอ็มเอ็นพี เป็นโปรโตคอลสำหรับใช้ตรวจสอบระดับการใช้งานทรัพยากรของอุปกรณ์ภายในระบบ ทำหน้าที่เก็บรวบรวมค่าการใช้งานทรัพยากรต่าง ๆ และนำข้อมูลที่ได้จัดเก็บลงในตารางที่ได้ออกแบบโครงสร้างไว้แล้วในฐานข้อมูล ซึ่งข้อมูลไม่ว่าจะมาจากเซิร์ฟเวอร์ประเภทไหนจะมีการจัดเก็บลงในตารางเหมือนกันโดยรูปที่แสดงด้านล่างจะเป็นการจัดเก็บข้อมูลการใช้งานทรัพยากรในแต่ละประเภท

id	source	date	time	cpu	cpu_level
1	CentLog	2013-03-08	13:56:57	0	1
2	CentLog	2013-03-08	13:57:08	0	1
3	CentLog	2013-03-08	13:57:18	0	1
4	CentLog	2013-03-08	13:57:28	0	1
5	CentLog	2013-03-08	13:57:38	1	1
6	CentLog	2013-03-08	13:57:48	1	1
7	CentLog	2013-03-08	13:57:59	2	1
8	CentLog	2013-03-08	13:58:09	2	1
9	CentLog	2013-03-08	13:58:19	2	1
10	CentLog	2013-03-08	13:58:29	2	1
11	CentLog	2013-03-08	13:58:39	1	1
12	CentLog	2013-03-08	13:58:49	5	1
13	CentLog	2013-03-08	13:58:59	11	1
14	CentLog	2013-03-08	13:59:09	10	1
15	CentLog	2013-03-08	13:59:19	10	1

รูปที่ 4.12 ข้อมูลการใช้งานหน่วยประมวลผล

id	source	date	time	mem	mem_level
1	CentLog	2013-03-08	13:57:08	19	1
2	CentLog	2013-03-08	13:57:18	19	1
3	CentLog	2013-03-08	13:57:28	45	3
4	CentLog	2013-03-08	13:57:38	78	4
5	CentLog	2013-03-08	13:57:48	98	5
6	CentLog	2013-03-08	13:57:59	98	5
7	CentLog	2013-03-08	13:58:09	98	5
8	CentLog	2013-03-08	13:58:19	98	5
9	CentLog	2013-03-08	13:58:29	97	5
10	CentLog	2013-03-08	13:58:39	97	5
11	CentLog	2013-03-08	13:58:49	97	5
12	CentLog	2013-03-08	13:58:59	97	5
13	CentLog	2013-03-08	13:59:09	97	5
14	CentLog	2013-03-08	13:59:19	97	5
15	CentLog	2013-03-08	13:59:30	97	5

รูปที่ 4.13 ข้อมูลการใช้งานหน่วยความจำ

id	source	date	time	bandwidth	bw_level
1	CentLog	2013-03-08	13:56:52	121	1
2	CentLog	2013-03-08	13:57:07	156	1
3	CentLog	2013-03-08	13:57:22	218	1
4	CentLog	2013-03-08	13:57:38	132718	1
5	CentLog	2013-03-08	13:57:53	362606	1
6	CentLog	2013-03-08	13:58:08	172512	1
7	CentLog	2013-03-08	13:58:23	259257	1
8	CentLog	2013-03-08	13:58:38	86991	1
9	CentLog	2013-03-08	13:58:53	230	1
10	CentLog	2013-03-08	13:59:08	156	1
11	CentLog	2013-03-08	13:59:24	214	1
12	CentLog	2013-03-08	13:59:39	152	1

รูปที่ 4.14 ข้อมูลการใช้งานแบนด์วิดท์ขาเข้า

id	source	date	time	bandwidth	bw_level
1	CentLog	2013-03-08	13:56:52	41	1
2	CentLog	2013-03-08	13:57:07	55	1
3	CentLog	2013-03-08	13:57:22	71	1
4	CentLog	2013-03-08	13:57:38	6324842	3
5	CentLog	2013-03-08	13:57:53	17408884	5
6	CentLog	2013-03-08	13:58:08	8085163	4
7	CentLog	2013-03-08	13:58:23	12434492	5
8	CentLog	2013-03-08	13:58:38	4170724	2
9	CentLog	2013-03-08	13:58:53	82	1
10	CentLog	2013-03-08	13:59:08	62	1
11	CentLog	2013-03-08	13:59:24	74	1
12	CentLog	2013-03-08	13:59:39	50	1

รูปที่ 4.15 ข้อมูลการใช้งานแบนด์วิดท์ขาออก

4.2.3 การจัดเก็บรายละเอียดข้อมูลที่ใช้ในการวิเคราะห์ความสัมพันธ์

ข้อมูลที่ได้จากทั้งการทำงานและการใช้งานทรัพยากรระบบนั้น จำเป็นต้องมีการนำมาจัดเก็บลงในตารางอีกตารางหนึ่งเพื่อเป็นการเตรียมข้อมูลก่อนการนำไปวิเคราะห์หาความสัมพันธ์ของข้อมูลทั้ง 2 ชนิด ซึ่งได้ผลดังรูปที่ 4.16

id	date	time	prun	cpuLevel	memLevel	bwInLevel	bwOutLevel
1	2013-03-02	03:38:40	GET/index.html	1	1	1	1
2	2013-03-02	03:38:49	GET/index2.html	1	1	1	1
3	2013-03-02	03:38:51	GET/dbpeak.php	2	1	1	1
4	2013-03-02	03:39:01	GET/test.php	3	1	1	1
5	2013-03-02	03:39:02	GET/webpeak.php	3	1	1	1
6	2013-03-02	03:39:11	GET/normaldb.php	3	1	1	1
7	2013-03-02	03:42:01	GET/index.html	1	1	1	1
8	2013-03-02	03:44:33	GET/dbpeak.php	1	1	1	1
9	2013-03-02	03:44:42	GET/index2.html	1	1	1	1
10	2013-03-02	03:45:22	GET/index.html	1	1	1	1
11	2013-03-02	03:46:33	GET/test.php	1	1	1	1
12	2013-03-02	03:47:53	GET/webpeak.php	2	1	1	1
13	2013-03-02	03:48:22	GET/normaldb.php	3	1	1	1
14	2013-03-02	03:48:42	GET/index.html	2	1	1	1
15	2013-03-02	03:50:12	GET/dbpeak.php	1	1	1	1
16	2013-03-02	03:50:30	GET/index2.html	1	1	1	1
17	2013-03-02	03:52:03	GET/index.html	1	1	1	1
18	2013-03-02	03:54:03	GET/test.php	1	1	1	1
19	2013-03-02	03:55:23	GET/index.html	1	1	1	1
20	2013-03-02	03:55:50	GET/dbpeak.php	1	1	1	1

รูปที่ 4.16 ข้อมูลเพื่อนำไปใช้ในการวิเคราะห์ความสัมพันธ์

4.2.4 การจัดเก็บรายละเอียดการแสดงผลข้อมูล

หลังจากนำข้อมูลไปวิเคราะห์หาความสัมพันธ์เรียบร้อย ผลลัพธ์ที่ได้จากการวิเคราะห์นั้นได้ถูกนำมาจัดเก็บลงในตารางที่ได้ออกแบบโครงสร้างแล้วในฐานข้อมูล ดังรูปที่ 4.17 เพื่อนำไปแสดงผลลัพธ์ยังหน้าเว็บแอพพลิเคชันสำหรับผู้ดูแลระบบ

id	Log	cpuLevel	memLevel	bwInLevel	bwOutLevel	conf
1	GET/index.html	null	1	null	null	1
2	GET/index.html	null	1	1	null	1
3	GET/index.html	null	1	1	1	1
4	GET/index.html	null	1	1	1	1
5	GET/index.html	null	1	1	1	1
6	GET/index.html	null	1	1	1	1
7	GET/index.html	null	1	1	1	1
8	GET/index.html	null	1	1	1	1
9	GET/index.html	null	1	1	1	1
10	GET/index.html	null	1	1	1	1
11	GET/index.html	null	1	1	1	1
12	GET/index.html	null	1	1	1	1
13	GET/index.html	null	1	1	1	1
14	GET/index.html	null	1	1	1	1
15	GET/index.html	null	1	1	1	1
16	GET/index.html	null	1	1	1	1
17	GET/index.html	null	1	1	1	1
18	GET/index.html	null	1	1	1	1
19	GET/index.html	null	1	1	1	1
20	GET/index.html	null	1	1	1	1

รูปที่ 4.17 ข้อมูลผลลัพธ์ที่ได้จากการวิเคราะห์ความสัมพันธ์

4.3 การวิเคราะห์ข้อมูล

ในส่วนของการสร้างโปรแกรมวิเคราะห์ข้อมูล ทางกลุ่มผู้วิจัยได้พัฒนาโปรแกรมด้วยภาษาจาวา โดยการนำไลบรารีของโปรแกรมวีก้ามาใช้ในการวิเคราะห์ข้อมูล ซึ่งมีการนำเทคนิคการทำเหมืองข้อมูลเข้ามาช่วยในการวิเคราะห์ คือ กฎความสัมพันธ์ของข้อมูล

4.3.1 วิธีหาความสัมพันธ์ด้วยกฎความสัมพันธ์ของข้อมูล

จากตารางที่ได้เก็บรวบรวมค่าทั้งจากการทำงานและการใช้งานทรัพยากรระบบในรูปที่ 4.16 เมื่อนำมาวิเคราะห์ด้วยโปรแกรมที่พัฒนาขึ้นซึ่งเป็นการนำออไพออรีอัลกอริทึมเป็นอัลกอริทึมที่ใช้ในการวิเคราะห์หาความสัมพันธ์ของข้อมูล จะได้ผลลัพธ์เป็นกฎของความสัมพันธ์ดังรูปที่....

```
Best rules found:

 1. webPrun=GET/my.php 11 ==> Date=2013-03-03 11    conf:(1)
 2. memLevel=3 10 ==> Date=2013-03-03 10    conf:(1)
 3. bwLevel=2 10 ==> Date=2013-03-03 10    conf:(1)
 4. bwLevel=1 9 ==> Date=2013-03-03 9    conf:(1)
 5. memLevel=3 bwLevel=2 9 ==> Date=2013-03-03 9    conf:(1)
 6. cpuLevel=1 7 ==> Date=2013-03-03 7    conf:(1)
 7. cpuLevel=3 7 ==> Date=2013-03-03 7    conf:(1)
 8. cpuLevel=1 7 ==> bwLevel=1 7    conf:(1)
 9. cpuLevel=3 7 ==> memLevel=3 7    conf:(1)
10. cpuLevel=3 7 ==> bwLevel=2 7    conf:(1)
11. webPrun=GET/my.php memLevel=3 7 ==> Date=2013-03-03 7    conf:(1)
12. cpuLevel=1 bwLevel=1 7 ==> Date=2013-03-03 7    conf:(1)
13. Date=2013-03-03 cpuLevel=1 7 ==> bwLevel=1 7    conf:(1)
14. cpuLevel=1 7 ==> Date=2013-03-03 bwLevel=1 7    conf:(1)
15. cpuLevel=3 memLevel=3 7 ==> Date=2013-03-03 7    conf:(1)
16. Date=2013-03-03 cpuLevel=3 7 ==> memLevel=3 7    conf:(1)
17. cpuLevel=3 7 ==> Date=2013-03-03 memLevel=3 7    conf:(1)
18. cpuLevel=3 bwLevel=2 7 ==> Date=2013-03-03 7    conf:(1)
19. Date=2013-03-03 cpuLevel=3 7 ==> bwLevel=2 7    conf:(1)
20. cpuLevel=3 7 ==> Date=2013-03-03 bwLevel=2 7    conf:(1)
```

รูปที่ 4.18 ผลลัพธ์จากการทำการวิเคราะห์ด้วยออไพออรีอัลกอริทึม

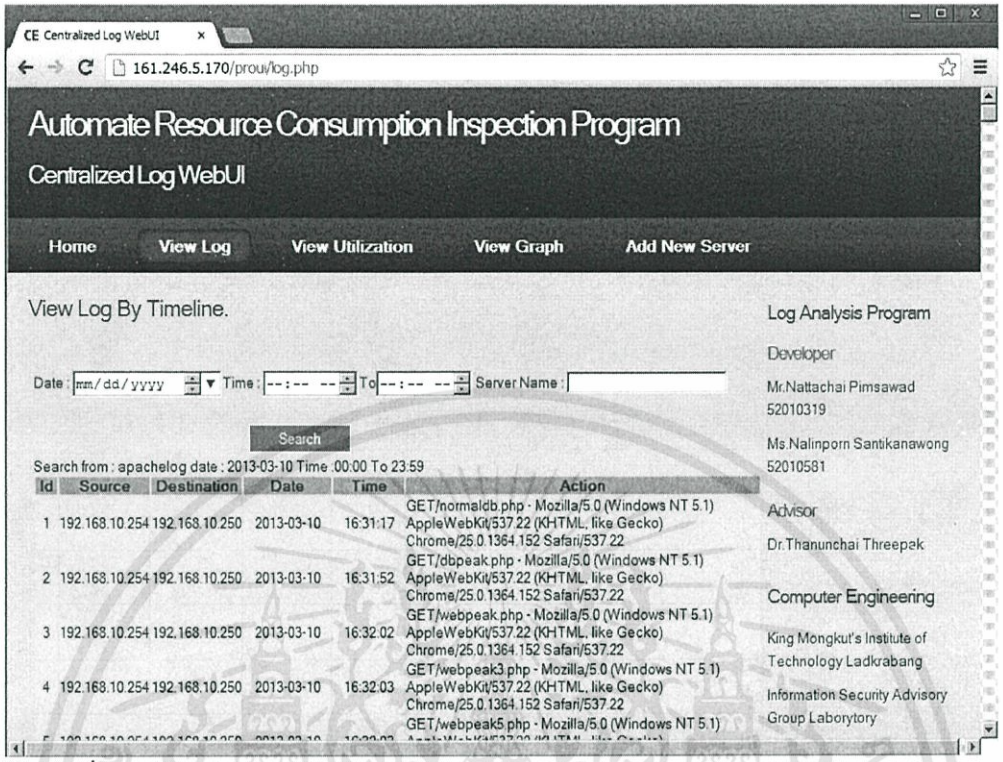
ผลลัพธ์จากรูปที่ 4.18 แสดงให้เห็นว่า ถ้าหากเกิดเหตุการณ์หนึ่งแล้วจะส่งผลให้เกิดอีกเหตุการณ์หนึ่ง โดยบางเหตุการณ์อย่างข้อมูลเรื่องวันที่จะไม่ได้ใช้ในการนำเสนอผลลัพธ์ด้วยกราฟความสัมพันธ์ ดังนั้นจึงต้องมีการคัดกรองกฎความสัมพันธ์เฉพาะบางกฎที่มีค่าทั้งการทำงานของระบบและการใช้งานของระบบ และนำค่าจากกฎความสัมพันธ์ดังกล่าวจัดเก็บลงในตารางนำเสนอผลลัพธ์ในฐานข้อมูล

4.4 การแสดงผล

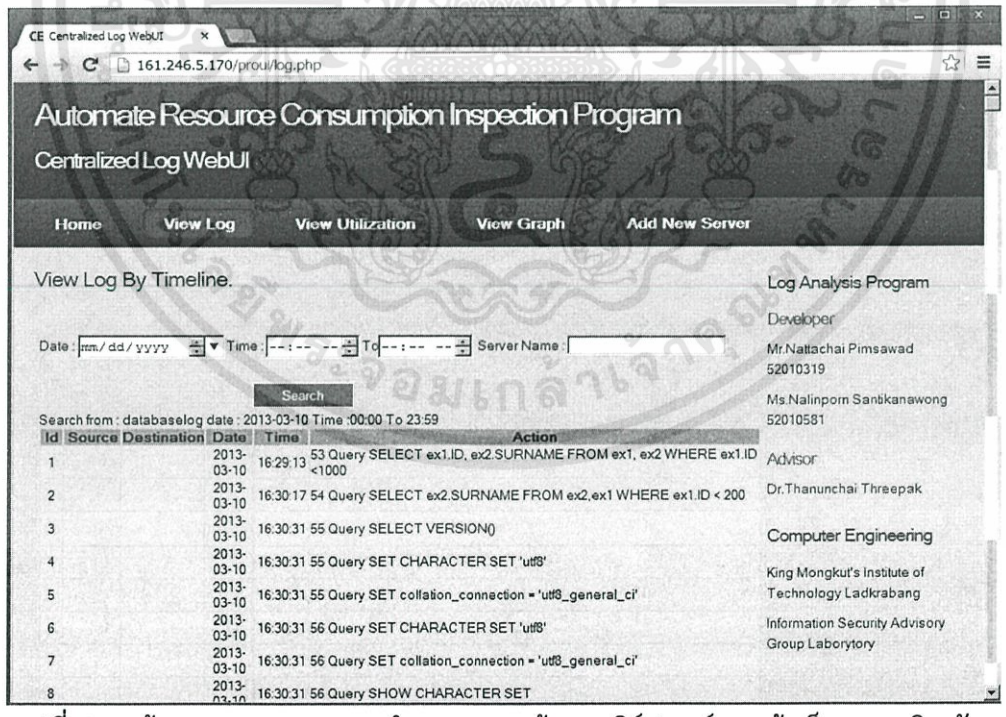
จากการเริ่มต้นตั้งแต่เก็บรวบรวมข้อมูลการทำงานของระบบและการใช้งานทรัพยากรประเภทต่าง ๆ ลงในฐานข้อมูล แล้วนำข้อมูลที่ได้นั้นไปทำการวิเคราะห์ด้วยโปรแกรมที่ทางกลุ่มพัฒนาขึ้น ผลลัพธ์ที่ได้จากทั้งการรวบรวมข้อมูลและการแสดงผลลัพธ์จากการวิเคราะห์ข้อมูลถูกนำเสนอด้วยเว็บแอพพลิเคชันซึ่งแสดงดังรูปต่อไปนี้

1) ส่วนแสดงผลรายละเอียดการทำงานของระบบ

ในส่วนของการแสดงรายละเอียดการทำงานของระบบจะแบ่งการทำงานของระบบออกเป็น 2 ส่วน คือ ผลรายละเอียดการทำงานจากเว็บเซิร์ฟเวอร์และผลรายละเอียดการทำงานจากดาต้าเบสเซิร์ฟเวอร์ ซึ่งแสดงผลของข้อมูลได้ดังรูปที่ 4.19 และรูปที่ 4.20 ตามลำดับ



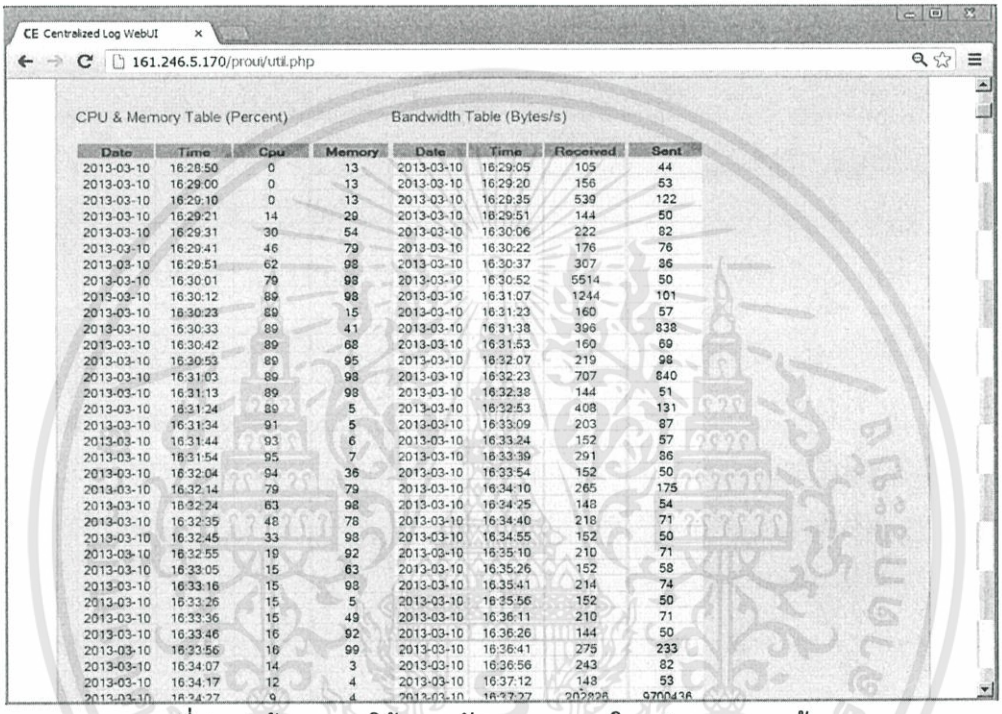
รูปที่ 4.19 ข้อมูลการแสดงผลการทำงานของเว็บเซิร์ฟเวอร์บนหน้าเว็บแอพพลิเคชัน



รูปที่ 4.20 ข้อมูลการแสดงผลการทำงานของดาต้าเบสเซิร์ฟเวอร์บนหน้าเว็บแอพพลิเคชัน

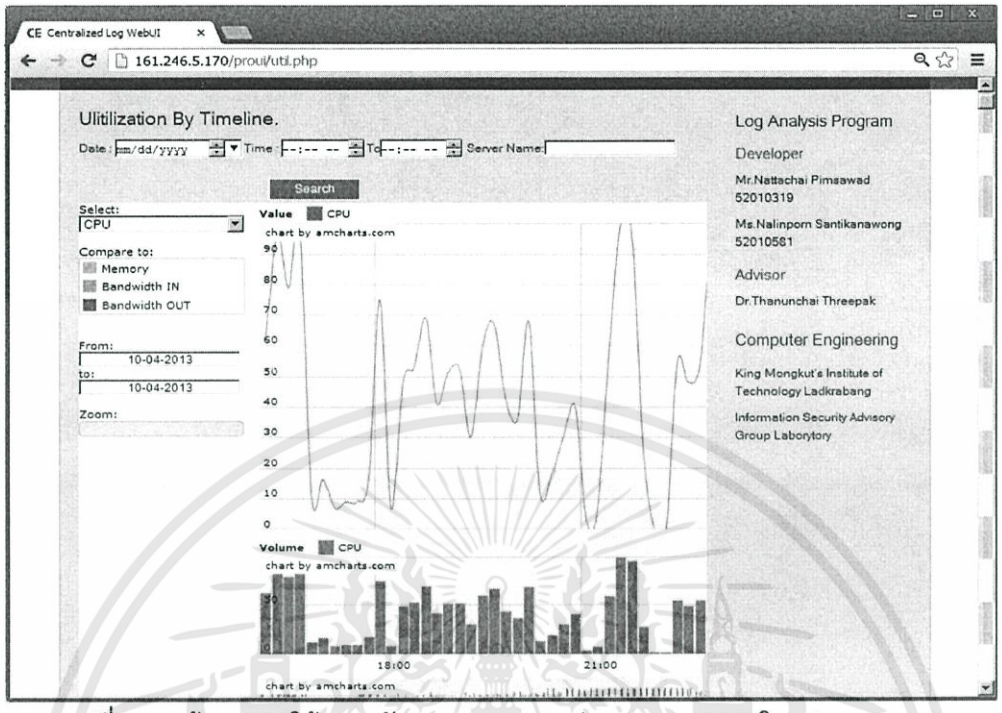
2) ส่วนแสดงผลการใช้ทรัพยากรระบบ

ในส่วนของข้อมูลการใช้ทรัพยากรระบบที่นำเสนอบนหน้าเว็บแอพพลิเคชันนั้นจะนำเสนอข้อมูลที่ได้ทำการรวบรวมทั้งจากหน่วยประมวลผล, หน่วยความจำ, แบนด์วิดท์ขาเข้า – ออก ตามรูปที่ 4.21 – 4.25 ตามลำดับ โดยเป็นการนำเสนอทั้งข้อมูลในรูปแบบตารางข้อมูลและกราฟ

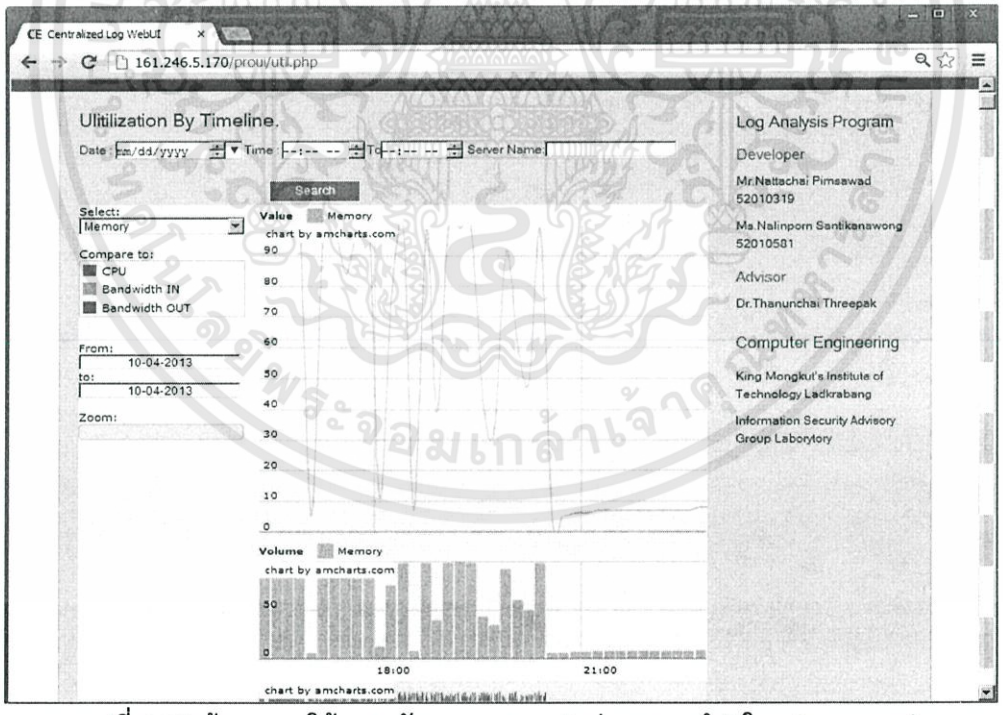


Date	Time	Cpu	Memory	Date	Time	Received	Sent
2013-03-10	16:28:50	0	13	2013-03-10	16:29:05	105	44
2013-03-10	16:29:00	0	13	2013-03-10	16:29:20	156	53
2013-03-10	16:29:10	0	13	2013-03-10	16:29:35	539	122
2013-03-10	16:29:21	14	29	2013-03-10	16:29:51	144	50
2013-03-10	16:29:31	30	54	2013-03-10	16:30:06	222	82
2013-03-10	16:29:41	46	79	2013-03-10	16:30:22	176	76
2013-03-10	16:29:51	62	98	2013-03-10	16:30:37	307	86
2013-03-10	16:30:01	79	98	2013-03-10	16:30:52	5514	50
2013-03-10	16:30:12	89	98	2013-03-10	16:31:07	1244	101
2013-03-10	16:30:23	89	15	2013-03-10	16:31:23	160	57
2013-03-10	16:30:33	89	41	2013-03-10	16:31:38	396	838
2013-03-10	16:30:42	89	68	2013-03-10	16:31:53	160	69
2013-03-10	16:30:53	89	95	2013-03-10	16:32:07	219	98
2013-03-10	16:31:03	89	98	2013-03-10	16:32:23	707	840
2013-03-10	16:31:13	89	98	2013-03-10	16:32:38	144	51
2013-03-10	16:31:24	89	5	2013-03-10	16:32:53	408	131
2013-03-10	16:31:34	91	5	2013-03-10	16:33:09	203	87
2013-03-10	16:31:44	93	6	2013-03-10	16:33:24	152	57
2013-03-10	16:31:54	95	7	2013-03-10	16:33:39	291	86
2013-03-10	16:32:04	94	36	2013-03-10	16:33:54	152	50
2013-03-10	16:32:14	79	79	2013-03-10	16:34:10	265	175
2013-03-10	16:32:24	63	98	2013-03-10	16:34:25	148	54
2013-03-10	16:32:35	48	78	2013-03-10	16:34:40	218	71
2013-03-10	16:32:45	33	98	2013-03-10	16:34:55	152	50
2013-03-10	16:32:55	19	92	2013-03-10	16:35:10	210	71
2013-03-10	16:33:05	15	63	2013-03-10	16:35:26	152	58
2013-03-10	16:33:16	15	98	2013-03-10	16:35:41	214	74
2013-03-10	16:33:26	15	5	2013-03-10	16:35:56	152	50
2013-03-10	16:33:36	15	49	2013-03-10	16:36:11	210	71
2013-03-10	16:33:46	16	92	2013-03-10	16:36:26	144	50
2013-03-10	16:33:56	16	99	2013-03-10	16:36:41	275	233
2013-03-10	16:34:07	14	3	2013-03-10	16:36:56	243	82
2013-03-10	16:34:17	12	4	2013-03-10	16:37:12	148	53
2013-03-10	16:34:27	9	4	2013-03-10	16:37:27	202826	9700436

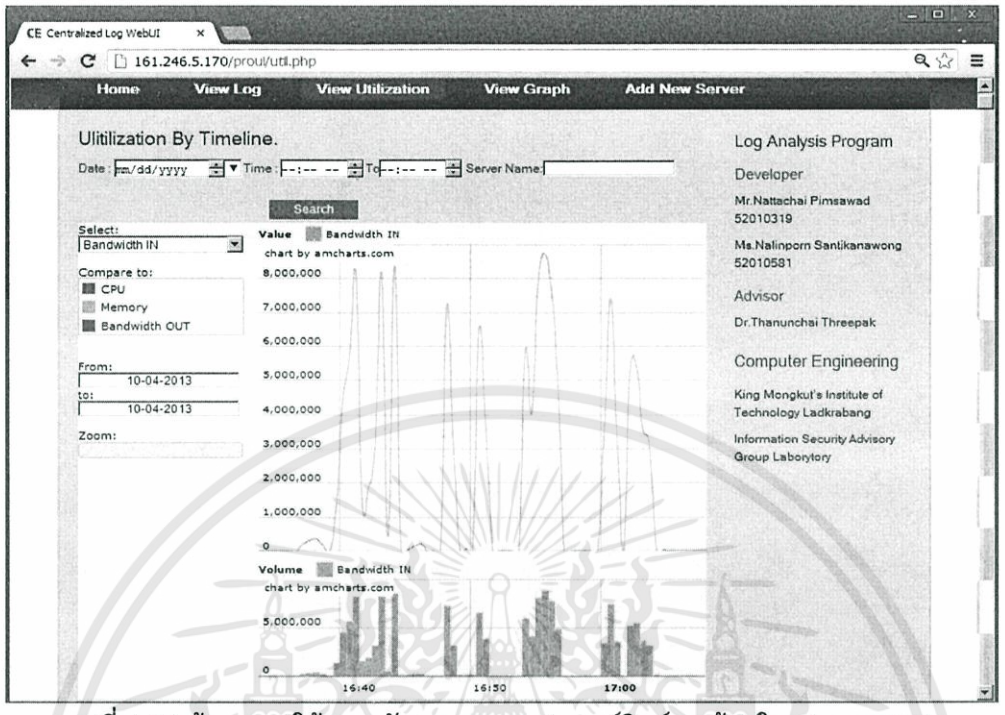
รูปที่ 4.21 ข้อมูลการใช้งานทรัพยากรระบบในรูปแบบตารางข้อมูล



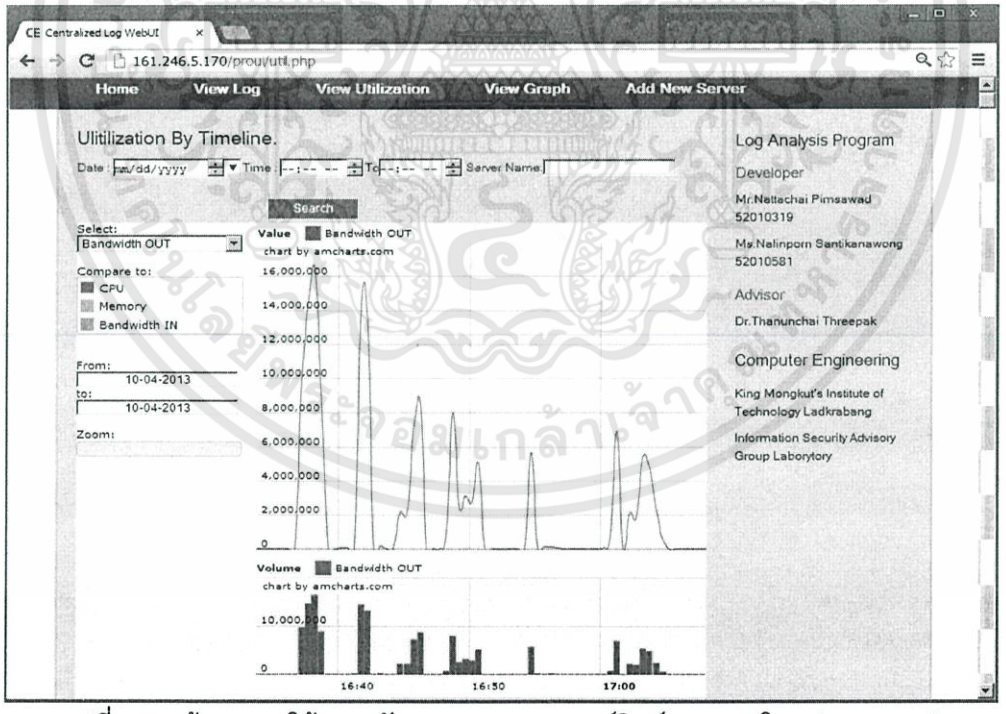
รูปที่ 4.22 ข้อมูลการใช้งานทรัพยากรระบบ (หน่วยประมวลผล) ในรูปแบบกราฟ



รูปที่ 4.23 ข้อมูลการใช้งานทรัพยากรระบบ (หน่วยความจำ) ในรูปแบบกราฟ



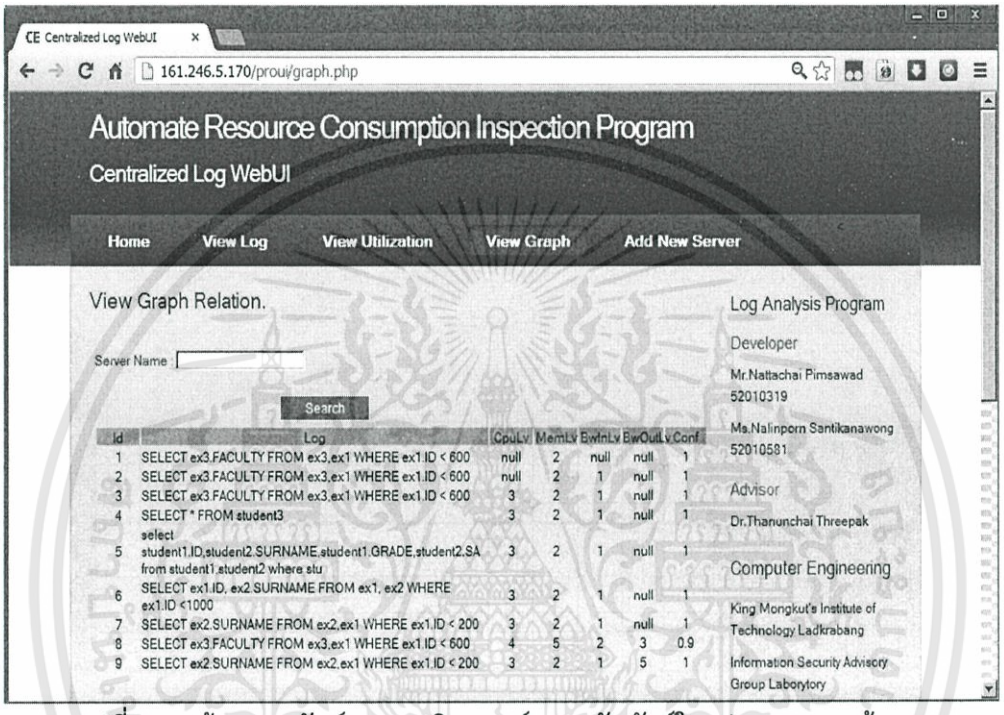
รูปที่ 4.24 ข้อมูลการใช้งานทรัพยากรระบบ (แบนด์วิดท์ขาเข้า) ในรูปแบบกราฟ



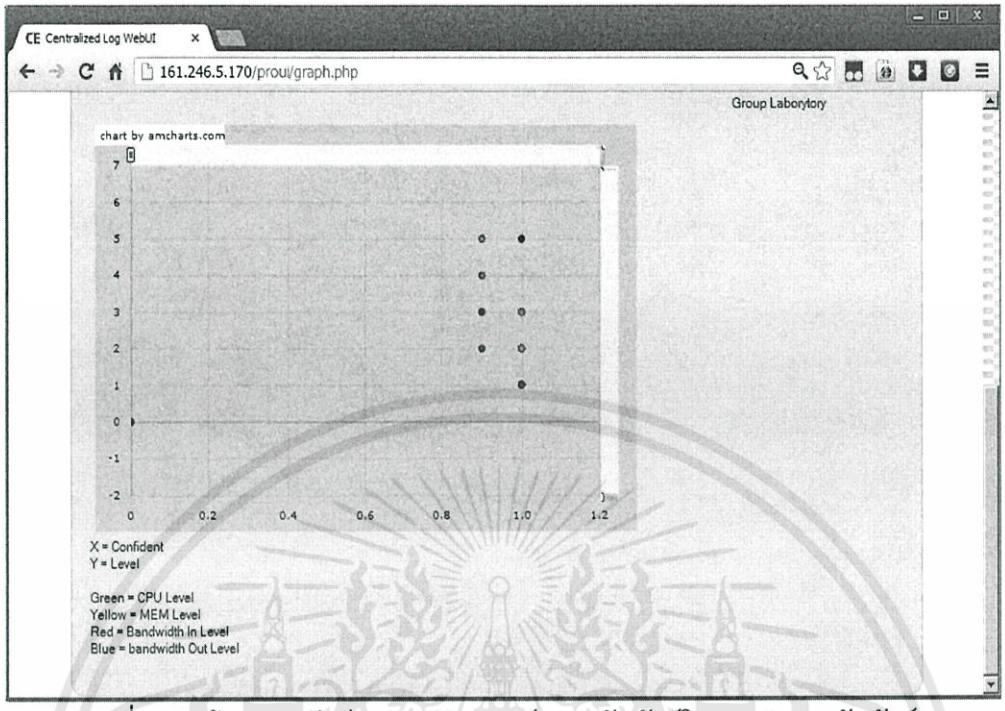
รูปที่ 4.25 ข้อมูลการใช้งานทรัพยากรระบบ (แบนด์วิดท์ขาออก) ในรูปแบบกราฟ

3) ส่วนแสดงผลลัพธ์ที่ได้จากการวิเคราะห์

ส่วนแสดงผลลัพธ์เป็นการนำผลลัพธ์ที่ได้จากการวิเคราะห์หาความสัมพันธ์ในรูปแบบของกราฟความสัมพันธ์ระหว่างค่าการใช้งานทรัพยากรระบบแต่ละประเภทกับค่าความเชื่อมั่นที่ได้จากกฎความสัมพันธ์ พร้อมทั้งตารางข้อมูลประกอบ



รูปที่ 4.26 ข้อมูลผลลัพธ์จากการวิเคราะห์ความสัมพันธ์ในรูปแบบตารางข้อมูล



รูปที่ 4.27 ข้อมูลผลลัพธ์จากการวิเคราะห์ความสัมพันธ์ในกราฟความสัมพันธ์

บทที่ 5
บทสรุปและข้อเสนอแนะ

5.1 บทสรุป

โครงงานโปรแกรมตรวจพินิจการใช้ทรัพยากรระบบสารสนเทศอัตโนมัติได้ทำการสร้างโปรแกรมที่สามารถใช้ในการตรวจสอบหาความสัมพันธ์ของข้อมูลการทำงานของระบบสารสนเทศและข้อมูลการใช้งานทรัพยากรประเภทต่าง ๆ ในระบบสารสนเทศนั้น ๆ เพื่อทำการตรวจสอบความผิดปกติที่อาจเกิดขึ้นในระบบสารสนเทศขนาดใหญ่จากการใช้งานทรัพยากรสารสนเทศที่มีมากเกินไป และหาสาเหตุได้ว่าเกิดจากการทำงานของระบบด้วยคำสั่งใด โดยทำการจำลองระบบเครือข่ายที่มีการเปิดให้บริการเว็บเซิร์ฟเวอร์และดาต้าเบสเซิร์ฟเวอร์ โดยทำการเก็บค่าการทำงานของระบบโดยติดตั้งซิสล็อก – เอ็นจี เพื่อใช้ในการรวบรวมข้อมูลดังกล่าว และในส่วนของการใช้งานทรัพยากรสารสนเทศนั้น ได้ทำการรวบรวมจากการใช้งานหน่วยประมวลผล หน่วยความจำ และแบนด์วิดท์ทั้งขาเข้า – ขาออก โดยทำการติดตั้งโปรโตคอลเอสเอ็นเอ็มพีเพื่อใช้ในการรวบรวมข้อมูลการใช้งานทรัพยากรระบบแล้วจึงนำข้อมูลทั้งสองส่วนที่ได้ทำการรวบรวมมาจัดเก็บลงในฐานข้อมูลที่เครื่องเซิร์ฟเวอร์กลาง และในส่วนของโปรแกรมที่ได้พัฒนาขึ้นนั้นจะเป็นการนำข้อมูลทั้งส่วนของการทำงานของระบบโดยใช้ล็อกการทำงานของแต่ละเซิร์ฟเวอร์ที่ได้ทำการกรองข้อมูลเรียบร้อยแล้วกับข้อมูลการใช้งานทรัพยากรแต่ละประเภทมาใช้ในการหาความสัมพันธ์ของข้อมูลดังกล่าว โดยการหากฎของความสัมพันธ์ด้วยออไพออรีอัลกอริทึม ซึ่งกฎของความสัมพันธ์ดังกล่าวจะคำนวณจากข้อมูลที่เกิดขึ้นบ่อย ๆ ในระบบและแสดงออกมาเป็นกฎของความสัมพันธ์ ผลลัพธ์ที่ได้จากการวิเคราะห์นี้ถูกนำเสนอด้วยเว็บแอพพลิเคชันในรูปแบบของตารางข้อมูล และกราฟความสัมพันธ์ระหว่างค่าความเชื่อมั่นของกฎความสัมพันธ์ และค่าการใช้งานทรัพยากรแต่ละประเภท ในการนำเสนอด้วยเว็บแอพพลิเคชันนั้นเพื่อสะดวกต่อการตรวจสอบความผิดปกติที่เกิดได้ง่ายของผู้ดูแลระบบ โดยผู้ดูแลระบบสามารถตรวจสอบได้จากกราฟแสดงความสัมพันธ์ ค่าความเชื่อมั่นที่มีค่าสูง ๆ แสดงว่าเป็นข้อมูลที่เชื่อถือได้ และหากค่าการใช้งานทรัพยากรประเภทใดที่มีค่าสูงด้วยแสดงว่า ล็อกการทำงานในตารางข้อมูลนั้นเป็นล็อกที่ก่อให้เกิดความผิดปกติขึ้นในระบบนั่นเอง

5.2 ปัญหาอุปสรรคและแนวทางการแก้ไข

เนื่องจากตอนแรก ข้อมูลที่จะใช้ในการวิเคราะห์ระบบจะเป็นการนำข้อมูลจริงจากสำนักทะเบียนและประมวลผล สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง มาใช้ในการ

ทดสอบโปรแกรมที่ได้พัฒนาขึ้น แต่เนื่องจากทางสำนักทะเบียนและประมวลได้มีการตั้งระบบภายในองค์กรใหม่ ทำให้มีปัญหาต่อการเข้าไปเก็บข้อมูล ดังนั้นจึงได้แก้ปัญหานี้ด้วยการที่ทางกลุ่มผู้วิจัยจึงได้จำลองระบบสารสนเทศขึ้นเอง และทำการเก็บข้อมูลจากระบบจำลองมาเพื่อใช้ในการทดสอบโปรแกรมที่พัฒนาขึ้น และการวิเคราะห์หาความสัมพันธ์ข้อมูล

5.3 แนวทางการพัฒนาต่อ

จากโครงงานนี้คาดว่าในอนาคตสามารถมีการพัฒนาต่อยอดโดยการนำแนวคิดในการเก็บรวบรวมข้อมูลและวิธีในการวิเคราะห์หาความสัมพันธ์ของข้อมูลใช้ในการพัฒนาจากการนำเสนอด้วยเว็บแอพพลิเคชันเป็นแอพพลิเคชันที่สามารถนำไปใช้ติดตั้งที่เครื่องคอมพิวเตอร์เพื่อใช้ในการดูแลจัดการระบบสารสนเทศ และสามารถใช้ในการตรวจสอบความผิดปกติต่าง ๆ ที่เกิดขึ้นในระบบสารสนเทศนั้น ๆ ได้สะดวกขึ้น

บรรณานุกรม

[1] ธนวัฒน์ แก้วบริสุทธิ์, นภดล ติรเดชาฤทธิ์. **"ระบบรวบรวมล็อกไฟล์ในสารบบและวิเคราะห์ข้อมูลโดยใช้เครื่องมือวิเคราะห์ธุรกิจอย่างชาญฉลาด"**ปริญญานิพนธ์สาขาวิชาวิศวกรรมศาสตร์คอมพิวเตอร์สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2554.

[2] ธันยพร ชาญพานิชกิจโชติ, นพรดา อริยนพรัตน์. **"การทำเหมืองข้อมูลบนกราฟิกโพรเซสเซอร์โดยใช้สถาปัตยกรรมคูด้า"** ปริญญานิพนธ์ สาขาวิชาวิศวกรรมศาสตร์คอมพิวเตอร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2554.

[3] บุญเสริม กิจศิริกุล.**"ปัญญาประดิษฐ์"** เอกสารคำสอนวิชา2110654, ภาควิชาวิศวกรรมคอมพิวเตอร์คณะวิศวกรรมคอมพิวเตอร์จุฬาลงกรณ์มหาวิทยาลัย, เวอร์ชัน 1.0.2 (มีนาคม 2548) : 191.