

บทที่ 3

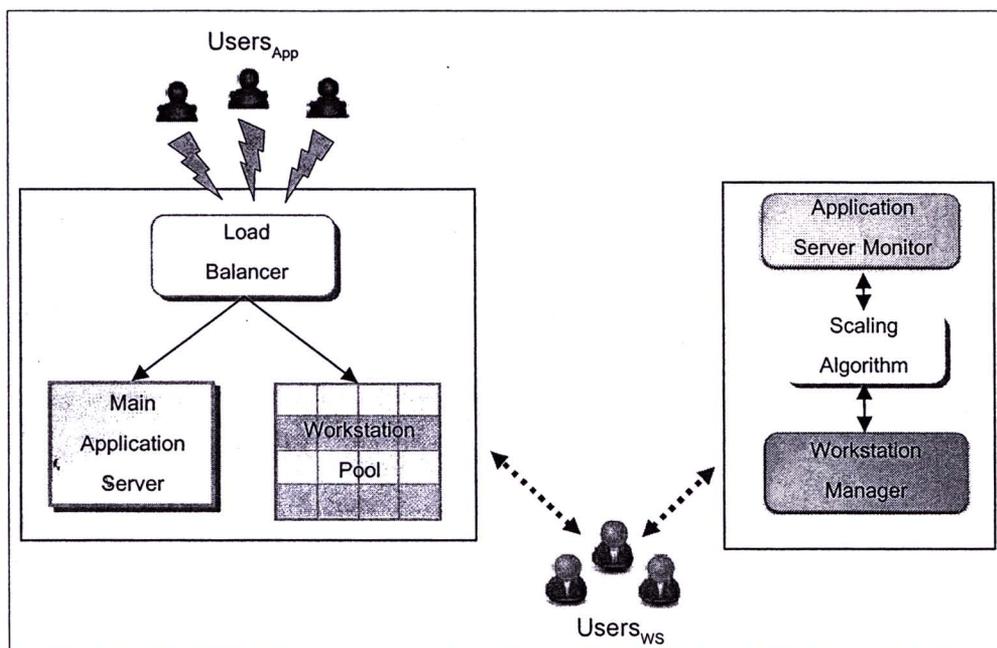
การจัดสรรทรัพยากรสำหรับโปรแกรมประยุกต์ผ่านเว็บโดยอาศัยทรัพยากรจากเครื่องเวิร์กสเตชัน

3.1 บทกล่าวนำ

งานวิจัยนี้มีแนวคิดในการนำทรัพยากรบางส่วนจากเครื่องเวิร์กสเตชันที่มีอยู่แล้วในหน่วยงาน และไม่ได้ถูกใช้งานอย่างเต็มที่มาใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชัน ในช่วงที่มีผู้ใช้งานเป็นจำนวนมาก เพื่อให้ระบบสามารถรองรับปริมาณการใช้งานจำนวนมาก เหล่านี้ให้ได้ ซึ่งจะช่วยให้ช่วยลดค่าใช้จ่ายในส่วน of เครื่องเซิร์ฟเวอร์สำรองลง รวมถึงจะช่วยให้เกิดการใช้ประโยชน์ของทรัพยากรที่ไม่ได้ถูกใช้งานอย่างเต็มที่ให้เกิดประโยชน์สูงขึ้นด้วย และงานวิจัยนี้ยังได้ทำการพัฒนาขั้นตอนวิธีในการขยายขนาดของเว็บแอปพลิเคชัน โดยปรับปรุงขั้นตอนวิธีการของ [4] จากเดิมซึ่งมีลักษณะการขยายขนาดได้ครั้งละเครื่อง ให้สามารถขยายขนาดของเว็บแอปพลิเคชันให้ได้ครั้งละหลายเครื่องตามปริมาณภาระงานที่เพิ่มขึ้นด้วย โดยเนื้อหาภายในบทนี้จะแบ่งออกเป็น 3 ส่วนคือ ในส่วนแรกจะเป็นบทกล่าวนำ ส่วนที่สองจะอธิบายเกี่ยวกับภาพรวมแนวความคิดด้านโครงสร้างสถาปัตยกรรมของระบบ และส่วนสุดท้ายจะเป็นการอธิบายถึงขั้นตอนวิธีการปรับขนาดของเว็บแอปพลิเคชันเซิร์ฟเวอร์ในระบบในลักษณะที่ละเครื่อง และขั้นตอนวิธีการปรับขนาดของเว็บแอปพลิเคชันเซิร์ฟเวอร์ในระบบในลักษณะที่ละหลายเครื่อง โดยใช้วิธีการคาดการณ์แนวโน้มของลักษณะภาระงานก่อนล่วงหน้า ตามลำดับ

3.2 ภาพรวมแนวความคิดด้านโครงสร้างสถาปัตยกรรม

ในหัวข้อนี้จะอธิบายสถาปัตยกรรมของระบบซึ่งประกอบไปด้วย 6 ส่วนหลักดังนี้ คือ 1) ส่วนกระจายภาระงาน 2) แอปพลิเคชันเซิร์ฟเวอร์หลักของระบบ 3) เครื่องเวิร์กสเตชันพลูในระบบ 4) ส่วนตรวจสอบสถานะของแต่ละแอปพลิเคชันเซิร์ฟเวอร์ 5) ขั้นตอนวิธีการปรับขนาดของเว็บแอปพลิเคชัน และ 6) ส่วนควบคุมการใช้งานของเครื่องเวิร์กสเตชันพลู ซึ่งภาพรวมของระบบที่นำเสนอได้แสดงดังรูปที่ 3.1



รูปที่ 3.1 แสดงภาพรวมของระบบที่นำเสนอ

3.2.1 ส่วนกระจายภาระงาน (Load Balancer)

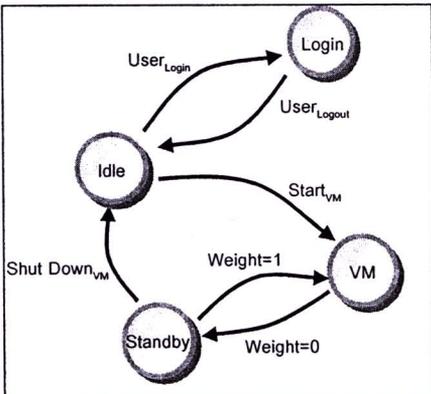
เป็นส่วนที่ทำหน้าที่ควบคุมการกระจายภาระงานจากผู้ใช้งานทั้งหมดโดยจะกระจายภาระงานในปริมาณที่เหมาะสมให้กับแต่ละเว็บแอปพลิเคชันเซิร์ฟเวอร์ที่กำลังเปิดให้บริการอยู่ในขณะนั้น สำหรับในกรณีที่มีปัญหาเกิดขึ้นกับเครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องใดหรือยังไม่พร้อมที่จะให้บริการ ส่วนกระจายภาระงานนี้ก็จะทำการปรับปรุงข้อมูลสถานะดังกล่าวแบบทันที (Real Time) เพื่อให้สามารถกระจายภาระงานให้กับเฉพาะเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะพร้อมที่จะให้บริการได้เท่านั้น

3.2.2 แอปพลิเคชันเซิร์ฟเวอร์หลักของระบบ (Main Application Server)

คือเครื่องแอปพลิเคชันเซิร์ฟเวอร์หลักที่ได้ถูกคิดให้กับระบบเว็บแอปพลิเคชัน ซึ่งจะสามารถรองรับปริมาณของผู้ใช้งานได้เฉพาะในช่วงปกติที่มีการใช้งานไม่มากเท่านั้น อย่างไรก็ตามสำหรับในช่วงที่มีผู้ใช้งานในระบบเว็บแอปพลิเคชันพร้อมกันเป็นจำนวนมาก (Peak Load) ทรัพยากรเครื่องเซิร์ฟเวอร์หลักนี้อาจไม่เพียงพอที่จะสามารถรองรับการให้บริการที่มีจำนวนมากได้

3.2.3 เครื่องเวิร์กสเตชันพลูในระบบ (Workstation Pool)

คือทรัพยากรเครื่องเวิร์กสเตชันทั้งหมดในระบบ ที่ไม่ได้ถูกใช้ประโยชน์อย่างเต็มที่ ซึ่งในงานวิจัยนี้ได้กำหนดให้ทรัพยากรเครื่องเวิร์กสเตชันในระบบนั้น มีเพียงพอสำหรับการให้บริการแก่ผู้ใช้งานเครื่องเวิร์กสเตชัน และมีเพียงพอต่อการใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันด้วย โดยที่ทรัพยากรเครื่องเวิร์กสเตชันเหล่านี้สามารถนำมาใช้จำลองเป็นเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือน เพื่อนำมาใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันในช่วงที่มีผู้ใช้งานพร้อมกันเป็นจำนวนมากได้ สำหรับในงานวิจัยนี้ได้ใช้เทคโนโลยีของเวอร์ชวลบ็อกซ์ไฮเปอร์ไวเซอร์ (Virtualbox Hypervisor) เวอร์ชัน 4.0 เพื่อใช้จำลองเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนให้กับระบบเว็บแอปพลิเคชัน เนื่องจากเทคโนโลยีของเวอร์ชวลบ็อกซ์ไฮเปอร์ไวเซอร์นั้นมีลักษณะเป็นโอเพ่นซอร์ส และมีความสามารถในการรองรับได้หลายระบบปฏิบัติการ เช่น ระบบปฏิบัติการวินโดวส์ (Windows), ระบบปฏิบัติการลินุกซ์ (Linux) หรือ ระบบปฏิบัติการสำหรับเครื่องคอมพิวเตอร์แมคอินทอช (Macintosh) เป็นต้น และจากลักษณะการใช้ประโยชน์จากทรัพยากรของเครื่องเวิร์กสเตชันดังที่ได้กล่าวมานั้น สามารถสรุปวงจรชีวิต (Life Cycle) ของแต่ละเครื่องเวิร์กสเตชันในระบบได้ตามรูปที่ 3.2 โดยที่แต่ละสถานะมีรายละเอียดดังนี้คือ



รูปที่ 3.2 แสดงวงจรชีวิตของเครื่องเวิร์กสเตชัน

สถานะ Login หมายถึง เครื่องเวิร์กสเตชันที่กำลังมีผู้ใช้งานลงบันทึกเข้า (Login) ใช้งานอยู่

Idle หมายถึง เครื่องเวิร์กสเตชันในเวิร์กสเตชันพูลที่ว่าง ไม่ได้ถูกใช้งาน

VM หมายถึง เครื่องเวิร์กสเตชันที่กำลังใช้เป็นที่พยากรณ์เสริมให้กับระบบเว็บแอปพลิเคชันในขณะนั้น

Standby หมายถึง เครื่องเวิร์กสเตชันที่กำลังใช้เป็นที่พยากรณ์เสริมให้กับเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนซึ่งกำลังจะถูกปิดการทำงาน (Shut Down) ลงในไม่ช้า โดยจะรอให้บริการกับการเชื่อมต่อที่ยังคงค้างอยู่ให้เสร็จสมบูรณ์ก่อน

3.2.4 ส่วนตรวจสอบสถานะของแต่ละแอปพลิเคชันเซิร์ฟเวอร์ (Application Server Monitor)

ในส่วนนี้จะทำหน้าที่ตรวจสอบและเก็บข้อมูลตัวชี้วัดการปรับขนาด (Scaling Indicator) ของระบบเว็บแอปพลิเคชันและข้อมูลสถานะของแต่ละแอปพลิเคชันเซิร์ฟเวอร์ เพื่อใช้เป็นข้อมูลสำหรับขั้นตอนวิธีในการปรับขนาดของระบบเว็บแอปพลิเคชันต่อไป

สำหรับตัวอย่างของตัวชี้วัดในการปรับขนาดของระบบเว็บแอปพลิเคชันที่นิยมใช้กัน ได้แก่ ปริมาณแอ็กทีฟคอนเนกชัน (Active Concurrent Connection) หรือค่าเฉลี่ยของระยะเวลาที่ระบบใช้ตอบสนองต่อผู้ใช้งาน (Average Response Time) เป็นต้น สำหรับในการใช้งานจริงนั้นก็ควรเลือกตัวชี้วัดให้เหมาะกับลักษณะของแต่ละระบบเว็บแอปพลิเคชันด้วย อย่างเช่น ในกรณีที่ระบบเว็บแอปพลิเคชันมีลักษณะเป็นแบบกระแสข้อมูล (Streaming) หรือระบบเว็บแอปพลิเคชันที่ต้องการขยายขนาดของแอปพลิเคชันเซิร์ฟเวอร์ให้ได้ครั้งละหลายเครื่อง การใช้ปริมาณแอ็กทีฟคอนเนกชันเป็นตัวชี้วัดในการสเกลนั้นอาจจะสะดวกกว่าการใช้ค่าเฉลี่ยของระยะเวลาที่ระบบใช้ตอบสนองต่อผู้ใช้งานเป็นตัวชี้วัดในการสเกล และเนื่องจากในงานวิจัยนี้ต้องการเพิ่มความสามารถให้ระบบเว็บแอปพลิเคชันสามารถขยายขนาดของแอปพลิเคชันเซิร์ฟเวอร์ให้ได้ครั้งละหลายเครื่องโดยการคาดการณ์แนวโน้มของปริมาณภาระงานก่อนล่วงหน้า ดังนั้นในเบื้องต้นจะทำการเลือกพิจารณาข้อมูลจำนวนแอ็กทีฟคอนเนกชันของแต่ละแอปพลิเคชันเซิร์ฟเวอร์นั้นได้ให้บริการ เพื่อใช้เป็นข้อมูลสำหรับขั้นตอนวิธีในการปรับขนาดของระบบเว็บแอปพลิเคชันต่อไป

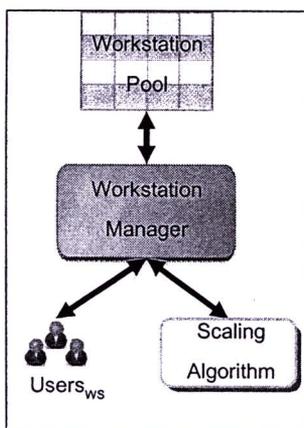
อย่างไรก็ตามในการใช้ปริมาณแอ็กทีฟคอนเนกชันเป็นตัวชี้วัดในการสเกลนั้นพบว่าอาจเกิดกรณีที่ลักษณะเนื้อหา (Content) ของแต่ละระบบเว็บแอปพลิเคชันนั้นอาจมีความหลากหลาย ซึ่งจะส่งผลให้แต่ละการเชื่อมต่อกันอาจมีปริมาณภาระงานที่แตกต่างกันได้ ดังนั้นในการใช้งานจริงควรมีการพิจารณาค่าน้ำหนัก (Weight) ของแต่ละประเภทของลักษณะภาระงานให้มีความสอดคล้องกับปริมาณค่าแอ็กทีฟคอนเนกชันหรือค่ามาตรฐาน (Baseline) เดียวกันก่อน เพื่อให้สามารถประเมินลักษณะปริมาณภาระงานโดยรวมของระบบเว็บแอปพลิเคชันดังกล่าวได้อย่างถูกต้องมากขึ้น

3.2.5 ขั้นตอนวิธีการปรับขนาดของระบบเว็บแอปพลิเคชัน (Scaling Algorithm)

ในส่วนของขั้นตอนวิธีการที่ใช้ในการปรับขนาดของระบบเว็บแอปพลิเคชันที่นำเสนอในงานวิจัยนี้มี 2 ลักษณะคือ ขั้นตอนวิธีการขยายขนาดของระบบเว็บแอปพลิเคชันในลักษณะที่ละเครื่อง ซึ่งโดยหลักการแล้วได้ใช้แนวคิดและวิธีการหลักจากงานวิจัย [4] คือในแต่ละครั้งที่ตรวจพบว่าควรที่จะขยายขนาดของระบบเว็บแอปพลิเคชัน เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชัน (Mng) จะทำการเพิ่มเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือน (Virtual Application Server) ให้แก่ระบบเพียงครั้งละหนึ่งเครื่องเท่านั้น อย่างไรก็ตามสำหรับในช่วงที่มีผู้ใช้งานเพิ่มขึ้นในปริมาณมาก มีความจำเป็นต้องใช้ทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนเพื่อรองรับปริมาณภาระงานในจำนวนหลายเครื่อง การเพิ่มปริมาณทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนในลักษณะที่ละเครื่องนี้อาจจะไม่เหมาะกับปริมาณภาระงานของระบบที่มีอัตราการเข้าใช้งานในลักษณะที่เพิ่มสูงขึ้นอย่างรวดเร็ว ดังนั้นผู้วิจัยจึงทำการพัฒนาขั้นตอนวิธีในการขยายขนาดของระบบเว็บแอปพลิเคชันดังกล่าว ให้สามารถขยายขนาดของระบบเว็บแอปพลิเคชันให้ได้ครั้งละหลายเครื่อง พร้อมกับเพิ่มความสามารถให้สามารถคาดการณ์แนวโน้มของปริมาณภาระงานในอนาคต ทำให้ระบบสามารถเตรียมทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนไว้ก่อนล่วงหน้าได้ จากลักษณะดังกล่าวช่วยทำให้ระบบสามารถรองรับปริมาณผู้ใช้งานที่มีอัตราการเข้าใช้งานในลักษณะเพิ่มสูงขึ้นอย่างรวดเร็วได้เป็นอย่างดี สำหรับขั้นตอนในการปรับขนาดของระบบเว็บแอปพลิเคชันทั้ง 2 ลักษณะนั้นได้อธิบายขั้นตอนวิธีการโดยละเอียดไว้ในหัวข้อที่ 3.3

3.2.6 ส่วนควบคุมการใช้งานของเครื่องเวิร์กสเตชันพูล (Workstation Pool Manager)

เนื่องจากในงานวิจัยนี้มีแนวคิดในการใช้ทรัพยากรบางส่วนจากเครื่องเวิร์กสเตชันที่มีอยู่แล้วในหน่วยงาน และไม่ได้ถูกใช้งานอย่างเต็มที่มาใช้เป็นทรัพยากรเสริมให้กับระบบ ดังนั้นจึงมีความจำเป็นที่จะต้องมียุทธศาสตร์ควบคุมการใช้งานของเครื่องเวิร์กสเตชันเหล่านี้ เพื่อให้สามารถถูกใช้ประโยชน์ได้อย่างเต็มที่และไม่รบกวนการใช้งานกัน โดยองค์ประกอบในส่วนนี้จะทำหน้าที่ตรวจสอบและเก็บสถานะของแต่ละเครื่องเวิร์กสเตชันในขณะนั้นว่าอยู่ในสถานะใดบ้าง หลังจากนั้นจะใช้ข้อมูลเหล่านี้ในการค้นหาเครื่องเวิร์กสเตชันที่เหมาะสมให้กับผู้มาขอใช้งานเครื่องเวิร์กสเตชัน หรือค้นหาทรัพยากรเครื่องเวิร์กสเตชันที่ไม่ได้ถูกใช้งานอย่างเต็มที่ เพื่อนำมาใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันในช่วงที่มีผู้ใช้งานพร้อมกันเป็นจำนวนมากได้โดยไม่ทำให้เกิดการแย่งการใช้งานทรัพยากรระหว่างกัน



รูปที่ 3.3 แสดงภาพรวมการทำงานในส่วนการควบคุมการใช้งานของเครื่องเวิร์กสเตชันพูล

สำหรับขั้นตอนวิธีในการควบคุมทรัพยากรในเครื่องเวิร์กสเตชันพูลนั้น เพื่อให้เกิดความเข้าใจได้ง่ายจะขออธิบายการทำงานออกเป็น 2 ส่วนหลักดังที่ได้แสดงในรูปที่ 3.3 คือ ขั้นตอนการทำงานที่ใช้ติดต่อกับผู้มาขอใช้งานเครื่องเวิร์กสเตชัน และขั้นตอนการทำงานที่ใช้ติดต่อกับส่วนของขั้นตอนวิธีในการปรับขนาดของระบบเว็บแอปพลิเคชันในหัวข้อที่ 3.2.6.1 และ 3.2.6.2 ตามลำดับ โดยในแต่ละขั้นตอนวิธีนั้นจะมีลักษณะการทำงานที่สอดคล้องกับวงจรชีวิตของแต่ละเครื่องเวิร์กสเตชันดังที่ได้แสดงไว้แล้วในรูปที่ 3.2

3.2.6.1 ขั้นตอนการทำงานที่ใช้ติดต่อกับผู้มาขอใช้งานเครื่อง เวิร์กสเตชัน

- 1) ในกรณีที่มีผู้ต้องการใช้งานเครื่องเวิร์กสเตชัน ผู้ใช้งานนั้นจะทำการขอหมายเลขเครื่องเวิร์กสเตชันจากในส่วนของควบคุมการใช้งานของเครื่องเวิร์กสเตชันพลุก่อน
- 2) ส่วนของควบคุมการใช้งานของเครื่องเวิร์กสเตชันพลุกจะทำการค้นหาเครื่องเวิร์กสเตชันที่มีสถานะ Idle คือไม่ได้ถูกใช้งานในขณะนั้น และส่งหมายเลขเครื่องดังกล่าวให้แก่ผู้ที่ต้องการใช้งาน
- 3) หลังจากนั้นส่วนของควบคุมการใช้งานของเครื่องเวิร์กสเตชันพลุก จะทำการเปลี่ยนสถานะของเครื่องเวิร์กสเตชันเครื่องนั้นจากเดิมมีสถานะ Idle ให้มีสถานะเป็น Login ซึ่งหมายความว่าผู้ใช้งานได้ทำการลงบันทึกเข้าใช้งานอยู่ในขณะนั้น
- 4) หลังจากที่ผู้ใช้งานเครื่องเวิร์กสเตชันดังกล่าวได้ใช้งานเสร็จเรียบร้อยแล้ว เมื่อผู้ใช้งานได้ทำการลงบันทึกออกจากระบบ (Logout) เครื่องเวิร์กสเตชันเครื่องนั้นก็ส่งข้อมูลกลับไปยังส่วนของควบคุมการใช้งานของเครื่องเวิร์กสเตชันพลุกว่าผู้ใช้งานได้คืนทรัพยากรให้กับระบบเรียบร้อยแล้ว
- 5) ส่วนของควบคุมการใช้งานของเครื่องเวิร์กสเตชันพลุกก็จะทำการเปลี่ยนสถานะของเครื่องเวิร์กสเตชันเครื่องนั้นจากสถานะ Login ให้มีสถานะเป็น Idle คือไม่ได้ถูกใช้งานและพร้อมที่จะนำทรัพยากรนี้ไปใช้งานอื่นได้อีกในครั้งต่อไป

3.2.6.2 ขั้นตอนการทำงานที่ใช้ติดต่อกับในส่วนควบคุมการปรับขนาดของระบบเว็บแอปพลิเคชัน

- 1) สำหรับในช่วงที่มีผู้ใช้งานในระบบเว็บแอปพลิเคชันพร้อมกันเป็นจำนวนมาก เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะตรวจพบได้ว่า ภาระงานในขณะนั้นมีปริมาณมากขึ้นส่งผลให้ทรัพยากรเว็บแอปพลิเคชันหลักในระบบมีไม่เพียงพอที่จะใช้รองรับปริมาณภาระงานดังกล่าว จำเป็นต้องเพิ่มจำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสริมให้แก่ระบบ ก็จะมาร้องขอทรัพยากรเครื่องเวิร์กสเตชันที่ไม่ได้ถูกใช้งานอย่างเต็มที่ เพื่อนำไปใช้จำลองเป็นทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือน เพื่อใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันในช่วงที่มีผู้ใช้งานพร้อมกันเป็นจำนวนมาก
- 2) ส่วนควบคุมการใช้งานของเครื่องเวิร์กสเตชันพลูจะทำการเลือกเครื่องเวิร์กสเตชันที่มีสถานะ Idle คือมีสถานะว่างและไม่ได้ถูกใช้งานอยู่ในขณะนั้น และส่งหมายเลขที่อยู่ไอพีของเครื่องเวิร์กสเตชันดังกล่าวให้กับส่วนที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชัน เพื่อนำไปใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันต่อไป
- 3) หลังจากนั้นส่วนควบคุมการใช้งานของเครื่องเวิร์กสเตชันพลู ก็ทำการเปลี่ยนสถานะของเครื่องเวิร์กสเตชันเครื่องนั้นจากเดิมมีสถานะเป็น Idle ให้มีสถานะเป็น VM ซึ่งหมายความว่ากำลังถูกใช้งานเป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันอยู่ในขณะนั้น
- 4) เมื่อผู้ใช้งานในระบบเว็บแอปพลิเคชันมีปริมาณลดลง ส่วนที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันก็จะตรวจพบได้ว่าภาระงานในขณะนั้นมีปริมาณน้อยลงด้วย และควรที่จะลดขนาดของทรัพยากรเว็บแอปพลิเคชันในระบบลง เพื่อให้สอดคล้องกับปริมาณการใช้งานในขณะนั้น ดังนั้นส่วนที่

ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันก็จะปิดการทำงาน (Shut Down) ของเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนลงเพื่อคืนทรัพยากรให้กับระบบ

- 5) หลังจากนั้นสถานะของเครื่องเวิร์กสเตชันดังกล่าวก็จะถูกเปลี่ยนจากเดิมที่มีสถานะเป็น VM ให้มีสถานะเป็น Standby และ Idle ตามลำดับและพร้อมที่จะนำทรัพยากรนี้ไปใช้งานอื่นได้อีกในครั้งต่อไป

ซึ่งโดยภาพรวมแล้วองค์ประกอบในส่วนนี้จะทำหน้าที่ควบคุมการใช้งานของทรัพยากรเครื่องเวิร์กสเตชันให้สามารถถูกใช้ประโยชน์ได้มากขึ้น ทั้งให้ใช้งานเป็นเครื่องเวิร์กสเตชันทั่วไปและใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันในช่วงที่มีการใช้งานเป็นจำนวนมากโดยไม่ให้มีการแย่งกันใช้งาน

3.3 ขั้นตอนวิธีการปรับขนาดของเว็บแอปพลิเคชันแบบไดนามิก (Dynamic Scaling Algorithm)

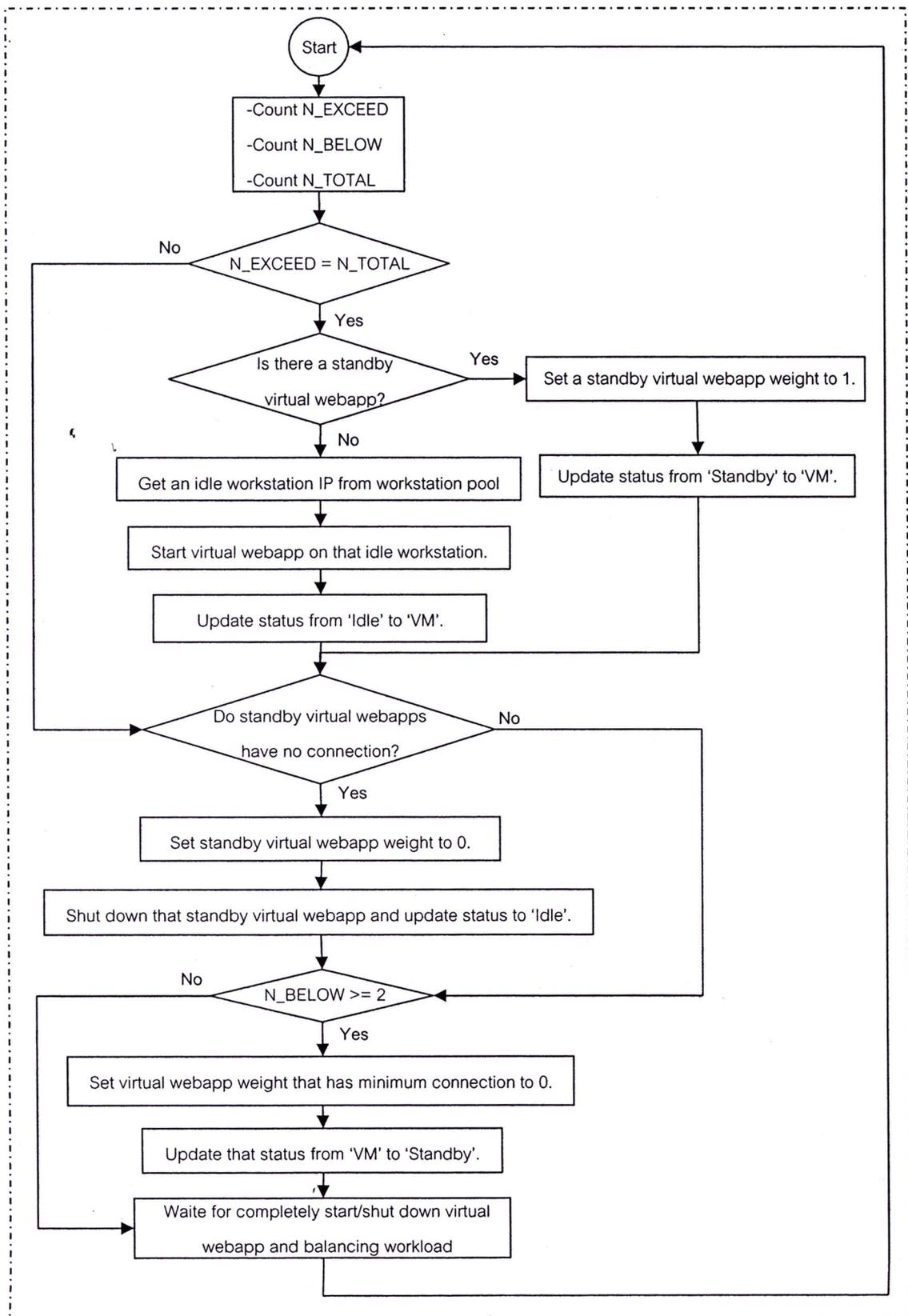
ในหัวข้อนี้จะกล่าวถึงรายละเอียดขั้นตอนวิธีการปรับขนาดของระบบเว็บแอปพลิเคชัน ซึ่งโดยภาพรวมแล้วก่อนที่จะทำการปรับขนาดของระบบเว็บแอปพลิเคชันจะต้องทำการกำหนดค่าขอบเขตบนสุด (Upper-threshold) และค่าขอบเขตล่างสุด (Lower-threshold) ที่จะใช้เป็นเกณฑ์ในการปรับขนาดของระบบเว็บแอปพลิเคชันก่อน โดยที่

ค่าขอบเขตบนสุด (Upper-threshold) คือ ค่าเปอร์เซ็นต์มากที่สุดของอัตราส่วนระหว่างจำนวนแอ็กทีฟคอนเน็กชันที่แต่ละแอปพลิเคชันเซิร์ฟเวอร์กำลังให้บริการอยู่ในขณะนั้น กับจำนวนแอ็กทีฟคอนเน็กชันที่แต่ละแอปพลิเคชันเซิร์ฟเวอร์เริ่มจะไม่สามารถรองรับได้ (ตัวอย่างเช่น กำหนดให้ค่าขอบเขตบนสุด มีค่าเท่ากับ 90% หมายความว่ากำหนดให้แต่ละแอปพลิเคชันเซิร์ฟเวอร์ในระบบ ควรมีปริมาณแอ็กทีฟคอนเน็กชันไม่เกิน 90% ของปริมาณแอ็กทีฟคอนเน็กชันที่แต่ละแอปพลิเคชันเซิร์ฟเวอร์เริ่มจะไม่สามารถรองรับได้)

ค่าขอบเขตล่างสุด (Lower-threshold) คือ ค่าเปอร์เซ็นต์น้อยที่สุดของอัตราส่วนระหว่างจำนวนแอ็กทีฟคอนเน็กชันที่แต่ละแอปพลิเคชันเซิร์ฟเวอร์กำลังให้บริการอยู่ในขณะนั้น กับจำนวนแอ็กทีฟคอนเน็กชันที่แต่ละแอปพลิเคชันเซิร์ฟเวอร์เริ่มจะไม่สามารถรองรับได้ (ตัวอย่างเช่น กำหนดให้ค่าขอบเขตล่างสุด มีค่าเท่ากับ 45% หมายความว่ากำหนดให้แต่ละแอป

พลิเคชันเซิร์ฟเวอร์ในระบบ ควรมีปริมาณแก็กที่พคอนเน็กชัน มากกว่า 45% ของปริมาณแก็กที่พคอนเน็กชันที่แต่ละแอปพลิเคชันเซิร์ฟเวอร์เริ่มจะไม่สามารถรองรับได้)

โดยหลักการแล้วเมื่อกำหนดค่าขอบเขตบนสุด และค่าขอบเขตล่างสุดที่จะใช้เป็นเกณฑ์ในการปรับขนาดของระบบเว็บแอปพลิเคชัน เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะทำการตรวจสอบจำนวนแก็กที่พคอนเน็กชันของแต่ละแอปพลิเคชันเซิร์ฟเวอร์ เปรียบเทียบกับค่าขอบเขตบนสุดและค่าขอบเขตล่างสุดที่ได้กำหนดไว้ เพื่อใช้ในการตัดสินใจในการขยายขนาดหรือลดขนาดของระบบเว็บแอปพลิเคชันตามปริมาณการใช้งานของผู้ใช้งานในขณะนั้น สำหรับขั้นตอนวิธีการปรับขนาดของระบบเว็บแอปพลิเคชันในลักษณะที่ละเครื่องและขั้นตอนวิธีการปรับขนาดของเว็บแอปพลิเคชันในลักษณะที่ละหลายเครื่องโดยใช้วิธีการคาดการณ์ล่วงหน้า สามารถเขียนเป็นแผนภาพแสดงขั้นตอนการทำงานได้ดังรูปที่ 3.4 และ 3.5 ตามลำดับ



รูปที่ 3.4 แสดงแผนภาพขั้นตอนวิธีการปรับขนาดของระบบเว็บแอปพลิเคชันในลักษณะที่ละเอียด

3.3.1 ขั้นตอนวิธีการปรับขนาดของระบบเว็บแอปพลิเคชันในลักษณะที่ ละเครื่อง (Dynamic Scaling Algorithm: Scale Each)

- 1) เริ่มต้นเครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะทำการตรวจสอบค่าต่างๆ ดังนี้
 - a. N_EXCEED คือ จำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในระบบที่มีจำนวน แอ็กทีฟคอนเนกชันเกินกว่าค่าขอบเขตบนสุด (Upper-threshold) ที่ได้กำหนดไว้ในขณะนั้น
 - b. N_BELOW คือจำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในระบบที่มีจำนวนแอ็กทีฟคอนเนกชันต่ำกว่าค่าขอบเขตล่างสุด (Lower-threshold) ที่ได้กำหนดไว้ในขณะนั้น
 - c. N_TOTAL คือ จำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์ทั้งหมดในระบบในขณะนั้น
- 2) เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะทำการตรวจสอบว่าจำนวน N_EXCEED ในขณะนั้นมีค่าเท่ากับจำนวน N_TOTAL หรือไม่
 - a. ในกรณีที่จำนวน N_EXCEED มีค่าไม่เท่ากับจำนวน N_TOTAL แสดงว่าระบบเว็บแอปพลิเคชันในปัจจุบันยังมีทรัพยากรเพียงพอที่จะใช้รองรับปริมาณภาระงานในขณะนั้น ดังนั้นเครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบจะยังไม่ทำการขยายขนาดของระบบเว็บแอปพลิเคชัน
 - b. ในกรณีที่จำนวน N_EXCEED มีค่าเท่ากับจำนวน N_TOTAL หมายความว่าทรัพยากรทั้งหมดในระบบเว็บแอปพลิเคชันเริ่มจะไม่สามารถรองรับปริมาณภาระงานในขณะนั้นได้ ดังนั้นเครื่องที่ควบคุมการขยายขนาดของระบบจะเริ่มทำการขยายขนาดของเว็บแอปพลิเคชัน โดยมีขั้นตอนดังนี้คือ

○ เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบจะทำการค้นหาว่าในขณะนั้นมีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby อยู่ในระบบบ้างหรือไม่

- ในกรณีที่มี

- เครื่องที่ควบคุมการขยายขนาดของระบบ จะทำการส่งข้อมูลไปยังเครื่องที่ทำหน้าที่กระจายภาระงานให้กับระบบเว็บแอปพลิเคชันให้ทำการเปลี่ยนแปลงค่าน้ำหนักของเครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องนั้น จากเดิมที่มีค่าน้ำหนักเท่ากับ 0 ให้มีค่าเท่ากับ 1 เพื่อให้สามารถใช้รองรับปริมาณภาระงานใหม่ที่จะเข้ามาในระบบได้ และทำการเปลี่ยนแปลงสถานะของเครื่องเวิร์กสเตชัน ที่ได้ใช้จำลองเป็นเครื่องแอปพลิเคชันเซิร์ฟเวอร์ดังกล่าวจากเดิมที่มีสถานะเป็น Standby ให้มีสถานะเป็น VM

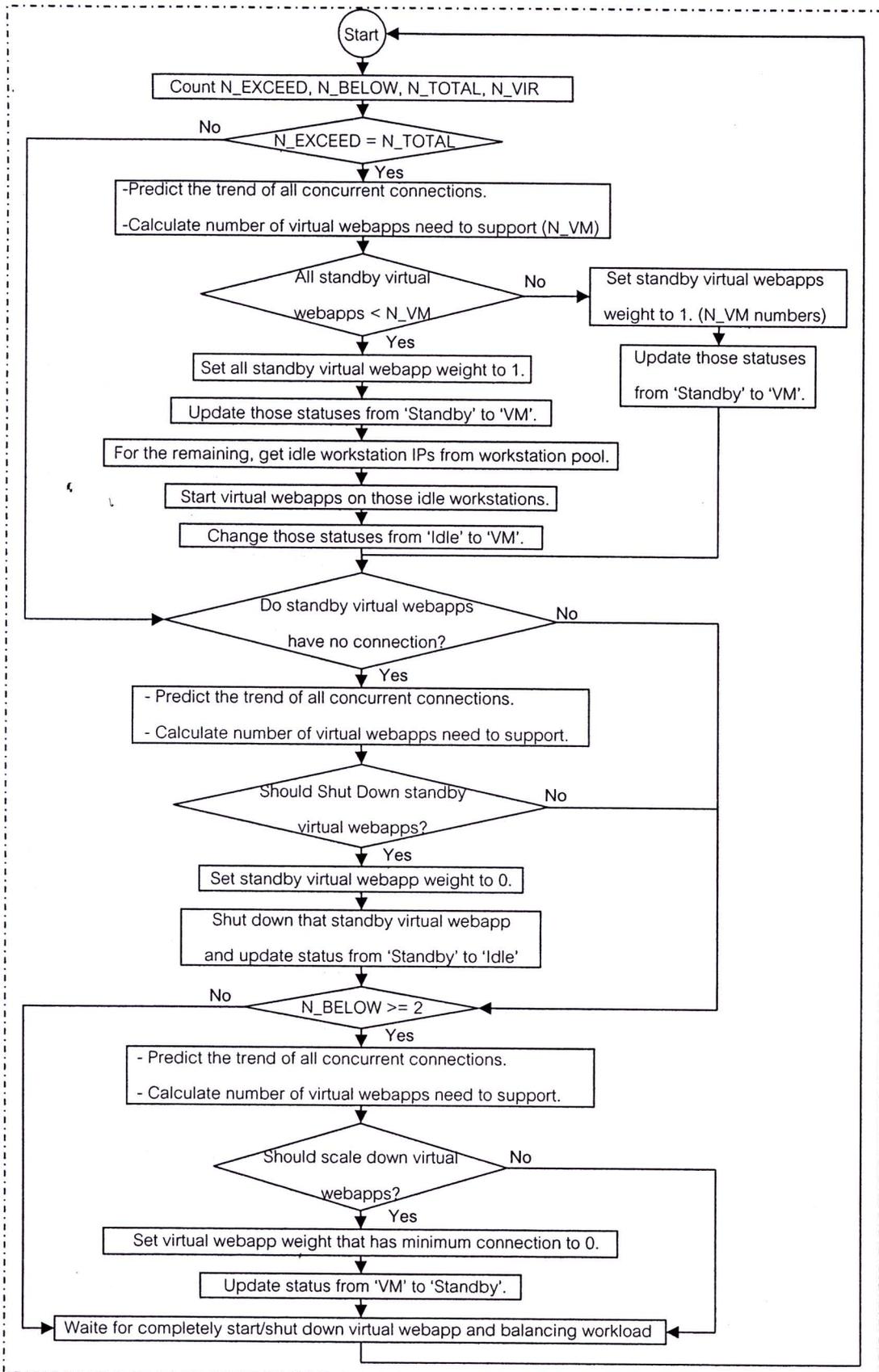
- ในกรณีที่ไม่มี

- เครื่องที่ควบคุมการขยายขนาดของระบบจะทำการค้นหาหมายเลขที่อยู่ไอพีของเครื่องเวิร์กสเตชันที่มีสถานะเป็น Idle ในเวิร์กสเตชันพลูจจำนวนหนึ่งเครื่อง

- หลังจากนั้นเครื่องที่ควบคุมการขยายขนาดของระบบ จะทำการจำลองเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนบนเครื่องเวิร์กสเตชันดังกล่าว เพื่อให้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชัน
 - เครื่องที่ควบคุมการขยายขนาดของระบบ จะทำการเปลี่ยนแปลงสถานะของเครื่องเวิร์กสเตชันที่ได้ใช้จำลองเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนดังกล่าว จากเดิมที่มีสถานะเป็น Idle ให้มีสถานะเป็น VM เพื่อแสดงว่ากำลังให้บริการแก่ผู้ใช้งานในระบบเว็บแอปพลิเคชัน
- 3) เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะทำการตรวจสอบว่าในขณะนั้นระบบมีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby และไม่มีการเชื่อมต่อกับผู้ใช้งานอยู่บ้างหรือไม่
- a. ในกรณีที่ไม่มีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในลักษณะดังกล่าว แสดงว่าในขณะนั้นระบบอาจไม่มีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะ Standby หรือเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะ Standby ในระบบกำลังให้บริการแก่ผู้ใช้งานอยู่ ซึ่งไม่ควรที่จะปิดการทำงานลง (Shut Down)
 - b. ในกรณีที่ระบบมีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในลักษณะดังกล่าว เครื่องที่ควบคุมการขยายขนาดของระบบจะปิดการทำงานของเครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องนั้นลงโดยมีขั้นตอนดังนี้คือ

- เครื่องที่ควบคุมการขยายขนาดของระบบจะส่งข้อมูลไปยังเครื่องที่ทำหน้าที่กระจายภาระงานให้กับระบบเว็บแอปพลิเคชัน ให้ทำการเปลี่ยนแปลงค่าน้ำหนักของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ดังกล่าวให้มีค่าเท่ากับศูนย์
 - หลังจากนั้นเครื่องที่ควบคุมการขยายขนาดของระบบจะสั่งให้เครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องดังกล่าวปิดการทำงานลง และเปลี่ยนแปลงสถานะของเครื่องเวิร์กสเตชันที่ได้ใช้จำลองเครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องนั้น จากเดิมที่มีสถานะ Standby ให้มีสถานะเป็น Idle แทน
- 4) เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะทำการตรวจสอบระบบว่า ในขณะนั้นมีจำนวน N_BELOW มากกว่าหรือเท่ากับสองเครื่องหรือไม่
- a. ในกรณีที่ค่าของ N_BELOW มีจำนวนไม่ถึงสองเครื่อง เครื่องที่ควบคุมการขยายขนาดของระบบจะยังไม่ทำการลดขนาดของระบบเว็บแอปพลิเคชันลง โดยถือว่าระบบเว็บแอปพลิเคชันในขณะนั้นยังมีการใช้งานในปริมาณที่มากอยู่ เนื่องจากยังมีการให้บริการของในแต่ละเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในปริมาณที่มากกว่าค่าขอบเขตล่างสุดที่ได้กำหนดไว้
 - b. ในกรณีที่ค่าของ N_BELOW มีจำนวนมากกว่าหรือเท่ากับ 2 เครื่อง เครื่องที่ควบคุมการขยายขนาดของระบบจะเริ่มทำการลดขนาดของระบบเว็บแอปพลิเคชันลง โดยมีขั้นตอนดังต่อไปนี้

- เครื่องที่ควบคุมการขยายขนาดของระบบจะส่งข้อมูลไปยังเครื่องที่ทำหน้าที่กระจายภาระงานให้กับระบบเว็บแอปพลิเคชัน ให้ทำการเปลี่ยนแปลงค่าน้ำหนักของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่ในขณะนั้นมีปริมาณการเชื่อมต่อกับผู้ใช้งานน้อยที่สุด ให้มีค่าน้ำหนักเท่ากับศูนย์ เพื่อที่จะได้ไม่ต้องส่งภาระงานใหม่ให้แก่เครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องดังกล่าว
 - เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบจะทำการเปลี่ยนแปลงสถานะของเครื่องเวิร์กสเตชันที่ได้ใช้จำลองเป็นเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนเครื่องดังกล่าว จากเดิมที่มีสถานะเป็น VM ให้มีสถานะเป็น Standby เพื่อแสดงว่ากำลังจะคืนทรัพยากรให้กับระบบ ทั้งนี้ยังไม่ทำการปิดการทำงานลง เนื่องจากได้เผื่อเวลาให้เครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องดังกล่าวสามารถให้บริการแก่การเชื่อมต่อที่อาจจะยังคงค้างการให้บริการอยู่
- 5) ระบบหยุดรอเวลาให้เครื่องแอปพลิเคชันเซิร์ฟเวอร์สามารถเปิดการให้บริการ (Start) หรือปิดการทำงานลง (Shut Down) ได้อย่างสมบูรณ์ รวมถึงรอเวลาให้ภาระงานได้กระจายให้กับแต่ละแอปพลิเคชันเซิร์ฟเวอร์ในระบบให้อยู่ในสภาวะสมดุลก่อนจะเริ่มทำงานในขั้นตอนที่ 1 ใหม่อีกครั้ง (เพื่อป้องกันการปรับขนาดของระบบเว็บแอปพลิเคชันในครั้งต่อไปที่อาจจะเกิดความผิดพลาดได้)



รูปที่ 3.5 แสดงแผนภาพขั้นตอนวิธีการปรับขนาดของระบบเว็บแอปพลิเคชันในลักษณะที่หลายเครื่องโดยใช้วิธีการคาดการณ์ล่วงหน้า

3.3.2 ขั้นตอนวิธีการปรับขนาดของระบบเว็บแอปพลิเคชันในลักษณะที่ หลายเครื่องโดยใช้วิธีการคาดการณ์ล่วงหน้า (Dynamic Scaling Algorithm: Scale Predict)

- 1) เริ่มต้นเครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะทำการตรวจสอบค่าต่างๆ ดังนี้
 - a. N_EXCEED คือ จำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในระบบที่มีจำนวน แอ็กทีฟคอนเน็กชันเกินกว่าค่าขอบเขตบนสุด (Upper-threshold) ที่ได้กำหนดไว้ในขณะนั้น
 - b. N_BELOW คือจำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในระบบที่มีจำนวน แอ็กทีฟคอนเน็กชันต่ำกว่าค่าขอบเขตล่างสุด (Lower-threshold) ที่ได้กำหนดไว้ในขณะนั้น
 - c. N_TOTAL คือ จำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์ทั้งหมดในระบบในขณะนั้น
 - d. N_VIR คือ จำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนที่ใช้เป็นทรัพยากรเสริมให้กับระบบและกำลังเปิดให้บริการอยู่ในขณะนั้น
- 2) เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะทำการตรวจสอบว่าจำนวน N_EXCEED ในขณะนั้นมีค่าเท่ากับจำนวน N_TOTAL หรือไม่
 - a. ในกรณีที่จำนวน N_EXCEED มีค่าไม่เท่ากับจำนวน N_TOTAL แสดงว่าระบบเว็บแอปพลิเคชันในปัจจุบันยังมีทรัพยากรเพียงพอที่จะใช้รองรับปริมาณภาระงานในขณะนั้น ดังนั้นเครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบจะยังไม่ทำการขยายขนาดของระบบเว็บแอปพลิเคชัน

- b. ในกรณีที่จำนวน N_EXCEED มีค่าเท่ากับจำนวน N_TOTAL หมายความว่าทรัพยากรทั้งหมดในระบบเว็บแอปพลิเคชันเริ่มจะไม่สามารถรองรับปริมาณภาระงานในขณะนั้นได้ ดังนั้น เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบจะเริ่มทำการขยายขนาดของระบบเว็บแอปพลิเคชัน โดยมีขั้นตอนดังนี้คือ
- เครื่องที่ควบคุมการขยายขนาดของระบบจะทำการคาดการณ์แนวโน้มของปริมาณแอ็กทีฟคอนเน็กชันที่ระบบจะต้องให้บริการในอนาคต โดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายในการวิเคราะห์
 - หลังจากนั้นระบบจะคำนวณจำนวนทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือน ว่าจะต้องเพิ่มจำนวนอีกกี่เครื่องเพื่อใช้รองรับปริมาณแอ็กทีฟคอนเน็กชันจากที่ได้คาดการณ์ไว้แล้ว ในที่นี้ใช้เป็นค่า N_VM เครื่อง
 - เครื่องที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะทำการเปรียบเทียบว่า ในปัจจุบันระบบมีปริมาณเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby น้อยกว่าปริมาณ N_VM ที่ได้จากการคำนวณหรือไม่
 - ในกรณีที่ปริมาณของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby มีปริมาณมากกว่าหรือเท่ากับค่า N_VM

- เครื่องที่ควบคุมการขยายขนาดของระบบ จะทำการส่งข้อมูลไปยังเครื่องที่ทำหน้าที่กระจายภาระงานให้กับระบบเว็บแอปพลิเคชัน ให้ทำการเปลี่ยนแปลงค่าน้ำหนักของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในระบบที่มีสถานะเป็น Standby ดังกล่าวจำนวน N_VM เครื่อง จากเดิมที่มีค่าน้ำหนักเท่ากับ 0 ให้มีค่าเท่ากับ 1 เพื่อให้สามารถรองรับปริมาณภาระงานใหม่ที่จะเข้ามาในระบบได้
 - เครื่องที่ควบคุมการขยายขนาดของระบบ จะทำการเปลี่ยนแปลงสถานะของเครื่องเวิร์กสเตชันที่ได้ใช้จำลองเป็นเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนดังกล่าว จากเดิมที่มีสถานะเป็น Standby ให้มีสถานะเป็น VM
- ในกรณีที่ปริมาณของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby มีปริมาณน้อยกว่าค่า N_VM
 - เครื่องที่ควบคุมการขยายขนาดของระบบจะทำการส่งข้อมูลไปยังเครื่องที่ทำหน้าที่กระจายภาระงานให้กับระบบเว็บแอปพลิเคชัน ให้ทำการเปลี่ยนแปลงค่าน้ำหนักของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby ทั้งหมดในระบบจากเดิมที่มีค่าน้ำหนักเท่ากับ 0 ให้มีค่าเท่ากับ 1

เพื่อให้สามารถใช้รองรับปริมาณภาระงานใหม่ที่จะเข้ามาในระบบได้

- และทำการเปลี่ยนแปลงสถานะของเครื่องเวิร์กสเตชันที่ได้ใช้จำลองเป็นเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนเหล่านั้น จากเดิมที่มีสถานะเป็น Standby ให้มีสถานะเป็น VM
- สำหรับทรัพยากรในส่วนที่เหลือ เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบ จะทำการค้นหาหมายเลขที่อยู่ไอพีของเครื่องเวิร์กสเตชันที่มีสถานะเป็น Idle ในเวิร์กสเตชันพูล จำนวน n เครื่อง (โดยที่ $n = N_{VM} -$ จำนวนเครื่องแอปพลิเคชันเซิร์ฟเวอร์ทั้งหมดที่มีสถานะเป็น Standby)
- หลังจากนั้นเครื่องที่ควบคุมการขยายขนาดของระบบ จะทำการจำลองเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนบนเครื่องเวิร์กสเตชันเหล่านั้นพร้อมกัน เพื่อใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชัน
- เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบ จะทำการเปลี่ยนสถานะของเครื่องเวิร์กสเตชันที่ได้ใช้จำลองเครื่องแอปพลิเคชันเสมือนเหล่านั้น จากเดิมที่มีสถานะเป็น Idle ให้มีสถานะเป็น VM เพื่อแสดงว่ากำลัง

ให้บริการแก่ผู้ใช้งานในระบบเว็บแอป
พลิเคชัน

- 3) เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบจะทำการตรวจสอบว่า ในขณะนั้นระบบมีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby และไม่มีการเชื่อมต่อกับผู้ใช้งานอยู่บ้างหรือไม่
- a. ในกรณีที่ไม่มีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในลักษณะดังกล่าว แสดงว่าในขณะนั้นระบบอาจไม่มีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะ Standby หรือเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะ Standby ในระบบกำลังให้บริการแก่ผู้ใช้งานอยู่ ซึ่งไม่ควรที่จะปิดการทำงานลง (Shut Down)
 - b. ในกรณีที่ระบบมีเครื่องแอปพลิเคชันเซิร์ฟเวอร์ในลักษณะดังกล่าว เครื่องที่ควบคุมการขยายขนาดของระบบจะปิดการทำงานของเครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องนั้นลงโดยมีขั้นตอนดังนี้คือ
 - เครื่องที่ควบคุมการขยายขนาดของระบบจะทำการคาดการณ์แนวโน้มของปริมาณแอ็กทีฟคอนเนกชันที่ระบบเว็บแอปพลิเคชันจะต้องให้บริการในอนาคต โดยใช้วิธีการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายในการวิเคราะห์ หลังจากนั้นระบบจะทำการคำนวณจำนวนทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่จะใช้สำหรับรองรับปริมาณแอ็กทีฟคอนเนกชันจากที่ได้คาดการณ์ไว้
 - จากข้อมูลดังกล่าว เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชัน จะสามารถตัดสินใจได้ว่าควรที่จะปิดการทำงานในส่วนหนึ่งของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby เหล่านี้หรือไม่

- ในกรณีที่ปริมาณภาระงานของระบบมีแนวโน้มเพิ่มสูงขึ้นและอาจจะต้องใช้ทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์ซึ่งในขณะนี้มีสถานะเป็น Standby เพื่อใช้เป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันอีกในอนาคต เครื่องที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันก็จะยังไม่ปิดการทำงานของเครื่องแอปพลิเคชันเซิร์ฟเวอร์เหล่านั้นลง
- ในกรณีที่ปริมาณภาระงานของระบบมีแนวโน้มลดลงหรือคงที่ ซึ่งระบบไม่จำเป็นต้องใช้ทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์เหล่านี้ มาใช้เป็นทรัพยากรเสริมอีกในอนาคต เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบก็จะปิดการทำงานของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่มีสถานะเป็น Standby เหล่านั้นลง โดยมีขั้นตอนดังต่อไปนี้ คือ
 - เครื่องที่ควบคุมการขยายขนาดของระบบ จะทำการส่งข้อมูลไปยังเครื่องที่ทำหน้าที่กระจายภาระงานให้กับระบบเว็บแอปพลิเคชันให้ทำการเปลี่ยนแปลงค่าน้ำหนักของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ดังกล่าวให้มีค่าเท่ากับศูนย์
 - หลังจากนั้นเครื่องที่ควบคุมการขยายขนาดของระบบจะสั่งให้เครื่องแอปพลิเคชันเซิร์ฟเวอร์ดังกล่าวปิดการทำงานลง และเปลี่ยนแปลงสถานะของเครื่องเวิร์กสเตชันที่ใช้จำลอง

เครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่อง
นั้น จากเดิมที่มีสถานะ Standby ให้มี
สถานะเป็น Idle แทน

- 4) เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันจะ
ทำการตรวจสอบระบบว่า ในขณะนั้นมีจำนวน N_BELOW มากกว่าหรือ
เท่ากับสองเครื่องหรือไม่
- a. ในกรณีที่ค่าของ N_BELOW มีจำนวนไม่ถึงสองเครื่อง เครื่องที่
ควบคุมการขยายขนาดของระบบจะยังไม่ทำการลดขนาดของ
ระบบเว็บแอปพลิเคชันลง โดยถือว่าระบบเว็บแอปพลิเคชันใน
ขณะนั้นยังมีการใช้งานในปริมาณที่มากอยู่ เนื่องจากยังมีการใช้
บริการของในแต่ละแอปพลิเคชันเซิร์ฟเวอร์ในปริมาณที่มากกว่า
ค่าขอบเขตล่างสุดที่ได้กำหนดไว้
 - b. ในกรณีที่ค่าของ N_BELOW มีจำนวนมากกว่าหรือเท่ากับ 2
เครื่อง เครื่องที่ควบคุมการขยายขนาดของระบบจะเริ่มทำการ
ลดขนาดของระบบเว็บแอปพลิเคชันลง โดยมีขั้นตอนดังต่อไปนี้
 - เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของ
ระบบ จะทำการคาดการณ์แนวโน้มของปริมาณแอ็กทีฟ
คอนเนกชัน ที่ระบบจะต้องให้บริการในอนาคต โดยใช้
วิธีการวิเคราะห์การถดถอยเชิงเส้นอย่างง่ายในการ
วิเคราะห์ หลังจากนั้นระบบจะทำการคำนวณจำนวน
ทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์ ที่จะใช้สำหรับ
รองรับปริมาณแอ็กทีฟคอนเนกชันจากที่ได้คาดการณ์
ไว้
 - จากข้อมูลดังกล่าว เครื่องที่ทำหน้าที่ควบคุม
การขยายขนาดของระบบเว็บแอปพลิเคชัน จะสามารถ
ตัดสินใจได้ว่าควรที่จะลดปริมาณเครื่องแอปพลิเคชัน
เซิร์ฟเวอร์ในระบบลงหรือไม่

- ในกรณีที่ปริมาณภาระงานของระบบมีแนวโน้มเพิ่มสูงขึ้นและอาจจะต้องใช้ทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์ดังกล่าวเป็นทรัพยากรเสริมให้กับระบบเว็บแอปพลิเคชันอีกในอนาคต เครื่องที่ควบคุมการขยายขนาดของระบบเว็บแอปพลิเคชันก็จะยังไม่ทำการลดขนาดของระบบเว็บแอปพลิเคชันลง
- ในกรณีที่ปริมาณภาระงานของระบบมีแนวโน้มลดลงหรือคงที่ ซึ่งระบบไม่จำเป็นต้องใช้ทรัพยากรเครื่องแอปพลิเคชันเซิร์ฟเวอร์เหล่านี้มาใช้เป็นทรัพยากรเสริมอีกในอนาคต เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบก็จะทำการลดขนาดของระบบเว็บแอปพลิเคชันลง โดยมีขั้นตอนดังต่อไปนี้ คือ
 - เครื่องที่ควบคุมการขยายขนาดของระบบจะทำการส่งข้อมูลไปยังเครื่องที่ทำหน้าที่กระจายภาระงานให้กับระบบเว็บแอปพลิเคชัน ให้ทำการเปลี่ยนแปลงค่าน้ำหนักของเครื่องแอปพลิเคชันเซิร์ฟเวอร์ที่ในขณะนั้นมีปริมาณการเชื่อมต่อกับผู้ใช้งานน้อยที่สุด ให้มีค่าน้ำหนักเท่ากับศูนย์ เพื่อที่จะได้ไม่ต้องส่งภาระงานใหม่ให้แก่เครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องดังกล่าว

- เครื่องที่ทำหน้าที่ควบคุมการขยายขนาดของระบบจะทำการเปลี่ยนแปลงสถานะของเครื่องเวิร์กสเตชันที่ได้ใช้จำลองเป็นเครื่องแอปพลิเคชันเซิร์ฟเวอร์เสมือนเครื่องดังกล่าว จากเดิมที่มีสถานะเป็น VM ให้มีสถานะเป็น Standby เพื่อแสดงว่ากำลังจะคืนทรัพยากรให้กลับระบบ ทั้งนี้ยังไม่ทำการปิดการทำงานลง เนื่องจากได้เผื่อเวลาให้เครื่องแอปพลิเคชันเซิร์ฟเวอร์เครื่องดังกล่าวสามารถให้บริการแก่การเชื่อมต่อที่อาจจะยังคงค้างการให้บริการอยู่

5) ระบบหยุดรอเวลาให้เครื่องแอปพลิเคชันเซิร์ฟเวอร์สามารถเปิดการให้บริการ (Start) หรือปิดการทำงานลง (Shut Down) ได้อย่างสมบูรณ์ รวมถึงรอเวลาให้ภาระงานได้กระจายให้กับแต่ละแอปพลิเคชันเซิร์ฟเวอร์ในระบบให้อยู่ในสภาวะสมดุลก่อนจะเริ่มทำงานในขั้นตอนที่ 1 ใหม่อีกครั้ง (เพื่อป้องกันการปรับขนาดของระบบเว็บแอปพลิเคชันในครั้งต่อไปที่อาจจะเกิดความผิดพลาดได้)