

# บทที่ 1

## บทนำ

### 1.1 ความเป็นมาและความสำคัญของปัญหา

จากอดีตจนถึงปัจจุบัน โลกของเรามีการพัฒนาทางด้านเทคโนโลยีไปอย่างมาก เริ่มตั้งแต่การเปลี่ยนแปลงจากยุคเกษตรกรรมไปสู่ยุคอุตสาหกรรม จนมาถึงปัจจุบัน ก็ยุคสารสนเทศ ซึ่งเป็นยุคที่ข้อมูลสารสนเทศเป็นสิ่งที่มีความจำเป็นกับชีวิตของมนุษย์อย่างขาดไม่ได้ และยังเป็นยุคที่มีการพัฒนาเทคโนโลยีสารสนเทศอย่างต่อเนื่อง ดังจะเห็นได้จาก การพัฒนาเทคโนโลยีคอมพิวเตอร์ ที่มีการผลิตไมโครโปรเซสเซอร์ (Microprocessor) ขึ้นมาใช้ ซึ่งทำให้คอมพิวเตอร์และอุปกรณ์มีขนาดเล็กลง แต่มีสมรรถนะสูงขึ้น มีการพัฒนาเทคโนโลยีเครือข่าย (Network Technology) ที่ช่วยให้สามารถติดต่อสื่อสารกันได้สะดวกและรวดเร็วขึ้น การพัฒนาเทคโนโลยีด้านการควบคุม ที่ทำให้เราสามารถควบคุมอุปกรณ์ต่างๆ ได้จากระยะไกล เป็นต้น ในขณะที่เทคโนโลยีสารสนเทศมีการเจริญก้าวหน้า มีการนำคอมพิวเตอร์เข้ามามีส่วนร่วมในการจัดการและจัดเก็บข้อมูล ซึ่งการจัดเก็บข้อมูลมีการจัดเก็บได้หลายรูปแบบ เช่น จัดเก็บในแฟ้มข้อมูล จัดเก็บในฐานข้อมูล โดยอุปกรณ์ที่ใช้ในการจัดเก็บข้อมูลอาจจะเป็น ฮาร์ดดิสก์ (Harddisk) หรือ เทป (Tape) แต่อุปกรณ์จัดเก็บข้อมูลของคอมพิวเตอร์ก็ยังมีข้อจำกัดระดับหนึ่ง รวมไปถึงแนวโน้มของข้อมูลสารสนเทศที่เพิ่มขึ้นอยู่ตลอดเวลา ส่งผลให้อุปกรณ์จัดเก็บข้อมูลไม่เพียงพอต่อความต้องการ วิธีการหนึ่งที่สามารถลดปัญหานี้ลงได้คือ การบีบอัดข้อมูล

การบีบอัดข้อมูล (Data Compression) [2] [7] [10] คือ การลดขนาดข้อมูลเพื่อให้ข้อมูลมีขนาดเล็กลง ทำให้สามารถเก็บข้อมูลได้มากขึ้นในขณะที่พื้นที่เก็บข้อมูลเท่าเดิม ซึ่งเทคโนโลยีนี้ถูกนำไปใช้งานในด้านต่างๆ เช่น การรักษาพื้นที่ว่างสำหรับดิสก์ การสำรองข้อมูล รวมถึงรูปภาพต่างๆ ที่อยู่ในเว็บเพจ จะถูกบีบอัดข้อมูล ซึ่งส่วนใหญ่อยู่ในรูปแบบ JPEG หรือ GIF อีกทั้งในหลายๆ ระบบแฟ้มข้อมูลจะมีการบีบอัดแฟ้มข้อมูลอัตโนมัติเมื่อมีการบันทึกข้อมูล เทคโนโลยีการบีบอัดข้อมูลนี้ไม่ได้ถูกใช้แค่เพียงแต่ในเครื่องคอมพิวเตอร์เท่านั้น แต่ยังถูกใช้ในโมเด็ม เราท์เตอร์ กล้องดิจิทัล โทรสาร และอื่นๆ อีกมากมาย ดังนั้น การศึกษาวิจัยเกี่ยวกับการบีบอัดข้อมูล จึงเป็นที่สนใจอย่างแพร่หลาย การบีบอัดข้อมูลด้วยวิธีบีบอัดเป็นการบีบอัดข้อมูลแบบอนุกรม (Sequential Compression) วิธีการหนึ่งที่ได้รับคามนิยมและมีประสิทธิภาพ อย่างไรก็ตามการบีบอัดข้อมูลแบบอนุกรมบนเครื่องคอมพิวเตอร์ที่มีหน่วยประมวลผลเดี่ยว (Sequential Computer) กับข้อมูลที่มีขนาดใหญ่ต้องใช้เวลาในการบีบอัดข้อมูลเป็นจำนวนมาก จึงเกิดแนวคิดที่จะนำเทคนิคของการประมวลผลแบบขนาน (Parallel Computing) บนเครื่องคอมพิวเตอร์ที่มีหลายหน่วยประมวลผล (Parallel Computer)[11] มาใช้ เพื่อลดเวลาในการบีบอัดข้อมูล

งานวิจัยนี้ได้นำเสนอการบีบอัดข้อมูลแบบขนานด้วยวิธีบีซีพทู (BZIP2) [3] ที่มีสมรรถนะ (Performance) มากขึ้นบนระบบพีซีคลัสเตอร์ (Cluster of PCs) ซึ่งการบีบอัดข้อมูลแบบขนานด้วยวิธีบีซีพทูนี้สามารถแบ่งเวลาที่ใช้ในการทำงานเป็น 2 ส่วน เช่นเดียวกับการประมวลผลแบบขนาน โดยทั่วไป คือ เวลาที่ใช้ในการประมวลผลแบบขนาน (Parallel Computation Time) และเวลาที่ใช้ในการติดต่อสื่อสารแบบขนาน (Parallel Communication Time) ในงานวิจัยนี้จะทำการเพิ่มสมรรถนะของเวลาที่ใช้ในการติดต่อสื่อสารแลกเปลี่ยนข้อมูลระหว่างหน่วยประมวล และลดการใช้งาน อุปกรณ์อินพุทเอาต์พุทศูนย์กลาง (Centralized I/O) พร้อมกันของแต่ละหน่วยประมวลผล โดยแสดงการเปรียบเทียบ ประเมินสมรรถนะจากเวลาที่ใช้ในการบีบอัดข้อมูล (Response Time) อัตราการเพิ่มขึ้นของความเร็ว (Speedup) และประสิทธิภาพ (Efficiency) ซึ่งจะพัฒนาโปรแกรมการบีบอัดข้อมูลด้วยภาษาซี (C Language) และมาตรฐานภาษาเอ็มพีไอ (MPI Standard) บนระบบพีซีคลัสเตอร์ เพื่อใช้ในการทดลอง และหาผลสรุปของงานวิจัย

## 1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

วิทยานิพนธ์ฉบับนี้มุ่งหวังเพื่อศึกษาการบีบอัดข้อมูลแบบขนานด้วยวิธีบีซีพทู (PBzip2) ที่มีผู้ศึกษาไว้บนเครื่องคอมพิวเตอร์แบบขนานที่ใช้หน่วยความจำร่วม (Shared-Memory Parallel Architectures) ซึ่งการบีบอัดข้อมูลแบบขนานด้วยวิธีบีซีพทูดังที่กล่าวมาแล้วข้างต้นเมื่อนำมาพัฒนาบนระบบพีซีคลัสเตอร์แล้วนั้นจะทำให้มีการติดต่อสื่อสารระหว่างหน่วยประมวลผลมาก ทำให้สมรรถนะของการบีบอัดข้อมูลต่ำลง ดังนั้นในวิทยานิพนธ์นี้จึงเสนอวิธีการลดการติดต่อสื่อสารแลกเปลี่ยนข้อมูลระหว่างหน่วยประมวลผล และวิธีการใช้งานอุปกรณ์อินพุทเอาต์พุทศูนย์กลางพร้อมกัน ของการบีบอัดข้อมูลแบบขนานด้วยวิธีบีซีพทู ซึ่งสามารถช่วยให้การบีบอัดข้อมูลมีสมรรถนะดีขึ้น

## 1.3 สมมติฐานของการศึกษา

- 1) การตัดหน่วยประมวลผลควบคุมออกเพื่อลดการติดต่อระหว่างหน่วยประมวลผลทำให้เวลาที่ใช้ในการติดต่อสื่อสารระหว่างหน่วยประมวลผลลดลง มีผลทำให้การบีบอัดข้อมูลแบบขนานด้วยวิธีบีซีพทูที่นำเสนอมีสมรรถนะเพิ่มขึ้น
- 2) การกำหนดน้ำหนักของการแบ่งบล็อกข้อมูลย่อย เพื่อให้บล็อกข้อมูลย่อยมีขนาดต่างกัน ทำให้แต่ละหน่วยประมวลผลใช้งานอุปกรณ์อินพุทเอาต์พุทศูนย์กลางไม่พร้อมกัน มีผลทำให้การบีบอัดข้อมูลแบบขนานด้วยวิธีบีซีพทูที่นำเสนอมีสมรรถนะเพิ่มขึ้น

## 1.4 ขอบเขตการวิจัย

ขอบเขตของวิทยานิพนธ์ฉบับนี้มีดังต่อไปนี้

- 1) การบีบอัดข้อมูลในงานวิจัยนี้ เป็นการบีบอัดข้อมูลแบบที่ไม่มีการสูญเสียข้อมูล (Lossless Compression)
- 2) งานวิจัยนี้ครอบคลุมเฉพาะส่วนการบีบอัดข้อมูลเท่านั้น จะไม่ครอบคลุมถึงส่วนของ การคลายข้อมูล
- 3) ข้อมูลที่ใช้สำหรับทดลองในงานวิจัยนี้ เป็นชุดข้อมูลที่นำมาจากชุดข้อมูลกูเทนเบิร์ก (Gutenberg Corpus)
- 4) ข้อมูลที่ได้จากการบีบอัดข้อมูลแบบขนานในงานวิจัยนี้ สามารถคลายข้อมูลกลับด้วยวิธีแบบอนุกรม
- 5) สมรรถนะของการบีบอัดข้อมูลวัดจากเวลาที่ใช้ในการบีบอัดข้อมูล อัตราการเพิ่มขึ้นของความเร็ว และประสิทธิภาพ
- 6) งานวิจัยนี้จะเปรียบเทียบสมรรถนะของการบีบอัดข้อมูลแบบขนานด้วยวิธีพีซีพทูแบบใหม่ที่เสนอ กับการบีบอัดข้อมูลแบบขนานด้วยวิธีพีซีพทูแบบเก่า ซึ่งพัฒนาบนระบบพีซีคลัสเตอร์

## 1.5 ขั้นตอนการศึกษาและดำเนินงานวิจัย

งานวิจัยนี้มีขั้นตอนการศึกษาและดำเนินงานวิจัยดังนี้

- 1) ศึกษางานวิจัยที่เกี่ยวข้องกับการบีบอัดข้อมูล
- 2) ศึกษาขั้นตอนวิธี (Algorithm) การบีบอัดข้อมูลด้วยวิธีพีซีพทู
- 3) ทำการตั้งสมมติฐาน ดังนี้ 1) ลดการติดต่อระหว่างหน่วยประมวลผลทำให้เวลาที่ใช้ในการติดต่อสื่อสารระหว่างหน่วยประมวลผลลดลง 2) ลดการใช้อินพุตเอาต์พุตศูนย์กลางพร้อมกัน โดยให้บล็อกข้อมูลย่อยมีขนาดต่างกัน ทำให้แต่ละหน่วยประมวลผลใช้งานอุปกรณ์อินพุตเอาต์พุตศูนย์กลางไม่พร้อมกัน มีผลทำให้การบีบอัดข้อมูลแบบขนานด้วยวิธีพีซีพทูที่นำเสนอมีสมรรถนะเพิ่มขึ้น
- 4) นำเสนอวิธีการบีบอัดข้อมูลแบบขนานด้วยวิธีพีซีพทูที่มีสมรรถนะเพิ่มขึ้นดังที่ได้ตั้งสมมติฐาน
- 5) พัฒนาโปรแกรมบีบอัดข้อมูลแบบขนานด้วยวิธีพีซีพทู โดยใช้ภาษาซีบนระบบพีซีคลัสเตอร์ที่ใช้ระบบปฏิบัติการลินุกซ์ (Linux) โดยใช้มาตรฐานภาษาเอ็มพีไอที่สนับสนุนการโปรแกรมแบบขนาน

- 6) วิเคราะห์ผลการทดลองโดยการเปรียบเทียบสมรรถนะของการบีบอัดข้อมูลแบบขนานด้วยวิธีบีบอัดนี้จากเวลาที่ใช้ในการบีบอัดข้อมูล อัตราการเพิ่มของความเร็ว และประสิทธิภาพ
- 7) สรุปผลการทดลองพร้อมเสนอแนวทางการพัฒนางานวิจัย
- 8) เขียนวิทยานิพนธ์

## 1.6 ประโยชน์ที่คาดว่าจะได้รับ

งานวิจัยนี้มีประโยชน์ที่คาดว่าจะได้รับจากการบีบอัดข้อมูลแบบขนานด้วยวิธีบีบอัดบนระบบพีซีคลัสเตอร์ ดังนี้

- 1) ลดเวลาที่ใช้ในการบีบอัดข้อมูล
- 2) นำมาใช้เป็นแนวทางในการพัฒนาระบบบีบอัดข้อมูลด้วยวิธีอื่น
- 3) นำมาใช้เป็นแนวทางในการพัฒนาโปรแกรมแบบขนานบนระบบพีซีคลัสเตอร์ เพื่อให้สามารถใช้ได้สะดวกและเข้าใจได้ง่ายมากขึ้น
- 4) นำมาใช้เป็นแนวทางในการพัฒนาระบบคอมพิวเตอร์แบบขนานชนิดคลัสเตอร์สำหรับหน่วยงานหรือสถานศึกษาที่ต้องการมีระบบคลัสเตอร์ของตนเอง