

หัวข้อวิทยานิพนธ์	การบีบอัดข้อมูลแบบขนานด้วยวิธีบีซิพทู บนระบบพีซีคลัสเตอร์
นักศึกษา	นายอัศวิน นิมกร
รหัสประจำตัว	47063707
ปริญญา	วิทยาศาสตรมหาบัณฑิต
สาขาวิชา	วิทยาการคอมพิวเตอร์
พ.ศ.	2549
อาจารย์ที่ปรึกษาวิทยานิพนธ์	ผศ. ดร. จีรพร ศรีสวัสดิ์

บทคัดย่อ

งานวิจัยนี้นำเสนอการบีบอัดข้อมูลแบบขนานด้วยวิธีบีซิพทูบนระบบพีซีคลัสเตอร์ทั้งหมด 3 วิธี คือ 1) วิธีเอ็มเอสบีซิพทู (MSBzip2) 2) วิธีเอดับเบิลยูบีซิพทู (AWBzip2) และ 3) วิธีเอดับเบิลยูดับเบิลยูบีซิพทู (AWWBzip2) โดยวิธีแรกเป็นวิธีที่ใช้แนวคิดการบีบอัดข้อมูลแบบขนานด้วยวิธีบีซิพทู (PBzip2) ที่มีผู้ศึกษาไว้บนเครื่องคอมพิวเตอร์แบบขนานที่ใช้หน่วยความจำร่วม (Shared-Memory Parallel Architectures) โดยมีหน่วยประมวลผลควบคุม ส่วนวิธีที่ 2 และ 3 เป็นวิธีที่เสนอแนวคิดใหม่โดยไม่มีหน่วยประมวลผลควบคุม เพื่อลดการติดต่อสื่อสารแลกเปลี่ยนข้อมูลระหว่างหน่วยประมวลผลควบคุมกับหน่วยประมวลผลบีบอัดข้อมูล ดังนั้นทุกหน่วยประมวลผลจะทำหน้าที่ แบ่งข้อมูล บีบอัดข้อมูล และเขียนข้อมูล ส่วนวิธีที่ 3 จะต่างจากวิธีที่ 2 ที่มีการกำหนดน้ำหนักของการแบ่งบล็อกข้อมูลย่อย เพื่อให้บล็อกข้อมูลย่อยมีขนาดต่างกัน ทำให้แต่ละหน่วยประมวลผลใช้งานอุปกรณ์อินพุทเอาต์พุทศูนย์กลาง (Centralized I/O) ไม่พร้อมกัน จากผลการทดลองบนระบบพีซีคลัสเตอร์ที่มีขนาด 1 ถึง 5 หน่วยประมวลผล พบว่า วิธีเอดับเบิลยูดับเบิลยูบีซิพทู (AWWBzip2) และวิธีเอดับเบิลยูบีซิพทู (AWBzip2) มีสมรรถนะดีกว่าวิธีเอ็มเอสบีซิพทู (MSBzip2) อย่างน้อย 19.91% และ 18.75% ตามลำดับ

Thesis	Parallel Data Compression with Bzip2 on a Cluster of PCs
Student	Mr.Assawin Nimkorn
Student ID	47063707
Degree	Master of Science
Program	Computer Science
Year	2006
Thesis Advisor	Asst. Prof. Dr. Jeeraporn Srisawat

ABSTRACT

This research proposes the parallel data compression with Bzip2 on a cluster of PCs. In particular, three strategies are introduced: 1) The MSBzip2 strategy, 2) The AWBzip2 strategy and 3) The AWWBzip2 strategy. The first method (MSBzip2) adapts the idea from the existing PBzip2, which is performed on shared-memory parallel architectures, controlled by the master processor. The last two methods (AWBzip2 and AWWBzip2) are presented with no master processor in order to reduce the communication time between the master processor and the slave processors. Therefore, in this case all processors will split the data into a number of blocks of the same size, compress the block of data and write the block of compressed data in parallel. The major difference of the last one (AWWBzip2) is that all processors will split the data into a number of blocks of the different weighted sizes in order to concurrent access the centralized input/output (I/O) device (or first-finish first-write). On the cluster system of size up to 5 processors (Ps), the experimental results showed that the AWWBzip2 strategy and the AWBzip2 strategy improved performance over the MSBzip2 strategy at least 19.91%.and 18.75%, respectively.