T 162964

บทคัดย่อ

การเรียงลำดับข้อมูลเป็นการประมวลผลขั้นพื้นฐานของงานด้านต่างๆ หลายด้านอาทิ งานด้านวิทยาศาสตร์และงานด้านวิศวกรรมศาสตร์ เป็นต้น ดังนั้นการศึกษาวิจัยเกี่ยวกับการเรียงลำดับข้อมูลจึงเป็นที่สนใจอย่างแพร่หลายทั้งด้านทฤษฎีและการประยุกต์ใช้ อย่างไรก็ตามการเรียงลำดับข้อมูลแบบอนุกรม ไม่สามารถตอบสนองความต้องการใช้ข้อมูลได้อย่างเต็มประสิทธิภาพ เนื่องจากปัจจุบันนี้ข้อมูลมีขนาดใหญ่ขึ้นและมีอัตราเพิ่มขึ้นอย่างรวดเร็ว จึงเกิดแนวความคิดที่จะนำการประมวลผลแบบขนานมาประยุกต์ใช้ในการเรียงลำดับข้อมูล งานวิจัยนี้ได้นำเสนอการเรียงลำดับแบบขนานด้วยวิธีไบโทนิค (Bitonic Sort) บนระบบพีซีคลัสเตอร์โดยใช้มาตรฐานภาษาเอ็มพีไอ ซึ่งการเรียงลำดับข้อมูลแบบขนานด้วยวิธีไบโทนิคนี้สามารถแบ่งเวลาที่ใช้ในการทำงานเป็น 2 ส่วน เช่นเดียวกับการประมวลผลแบบขนานโดยทั่วไป คือ เวลาที่ใช้ในการประมวลผลแบบขนานและเวลาที่ใช้ในการติดต่อสื่อสารแบบขนาน ในงานวิจัยนี้จะทำการเพิ่มประสิทธิภาพของเวลาที่ใช้ในติดต่อสื่อสารระหว่างหน่วยประมวลผลในการเรียงลำดับแบบขนานด้วยวิธีไบโทนิค และลดการใช้หน่วยความจำในการเก็บข้อมูลที่ซ้ำซ้อนกันในระบบ ซึ่งเสนอไว้ 4 วิธี คือ 1) วิธีบีเอส (BS) 2) วิธีซีอีบีเอส (CEBS) 3) วิธีเอสอีบีเอส (SEBS) 4) วิธีซีเอสอีบีเอส (CSEBS) โดยสองวิธีแรกเป็นวิธีที่มีผู้เสนอไว้แล้ว ส่วนสองวิธีหลังเป็นวิธีใหม่ที่เสนอขึ้นในวิทยานิพนธ์นี้ นอกจากนั้นทุกวิธีดังกล่าวถูกพัฒนาด้วยภาษาซีและมาตรฐานภาษาเอ็มพีไอบนระบบพีซีคลัสเตอร์เพื่อใช้ในการทดลอง เปรียบเทียบ และหาผลสรุปของงานวิจัย โดยแสดงการเปรียบเทียบวิธีที่เสนอและวิธีเดิมด้วยการประเมินผลจากเวลาที่ใช้ในการเรียงลำดับข้อมูล (Response Time) อัตราการเพิ่มขึ้นของความเร็ว (Speedup) และประสิทธิภาพ (Efficiency) จากผลการทดลองบนระบบพีซีคลัสเตอร์ที่มีขนาด 1 ถึง 8 หน่วยประมวลผล (P) พบว่าวิธีใหม่ที่เสนอซีเอสอีบีเอสมีค่าอัตราการเพิ่มขึ้นของความเร็วเข้าใกล้กรณีอุดมคติมาก คือ 1.9, 3.3 และ 7.8 ในกรณีที่จำนวนหน่วยประมวลผลเท่ากับ 2, 4 และ 8 หน่วยตามลำดับ และเมื่อเปรียบเทียบกับวิธีที่มีอยู่เดิม พบว่าวิธีซีเอสอีบีเอสที่เสนอมีประสิทธิภาพดีกว่าวิธีบีเอสอย่างน้อย 20% และดีกว่าวิธีซีอีบีเอสอย่างน้อย 5%

TE162964

ABSTRACT

Sorting is one of the most common processing performed by computers in many applications such as scientific and engineering applications. Sorting technique is very popular in both the theoretical study and its widely utilized applications. However sequential sorting on large data sets is very time consuming and cannot properly manage data, which have rapidly growth rate. Hence parallel sorting has been introduced for such sorting applications. The purpose of this research is to present "parallel Bitonic sorting" for the very large data sets on a cluster of PCs by using MPI standard. Like any parallel processing, time complexity of the parallel Bitonic sorting consists of both the parallel computation time and the parallel communication time. This study focuses on the improvement of the communication time of the existing parallel Bitonic sorting and also the efficient used of the memory overhead. In particular, 4 strategies are purposed: 1) the BS (Bitonic Sort) strategy, 2) the CEBS (Communication-Efficient Bitonic Sort) strategy, 3) the SEBS (Space-Efficient Bitonic Sort) strategy and 4) the CSEBS (Communication-Space Efficient Bitonic Sort) strategy. The first two methods are the existing ones, whereas the last two strategies are introduced in this thesis. In order to measure system performance, all four methods were developed by using C language and MPI standard on the cluster of PCs. In such cluster environment, the system performance of new parallel Bitonic sorting (SEBS and CSEBS strategies) and existing Bitonic sorting (BS and CEBS strategies) were compared in terms of response time, speedup, and efficiency. On the cluster system of size up to 8 processors (P), the experimental results showed that the CSEBS strategy yielded nearly ideal speedup, which were 1.9, 3.3 and 7.8 for P = 2, 4 and 8, respectively. In addition, the CSEBS method performed the improved performance at least 20% over that of the original BS method and at least 5% over that of the existing CEBS method.