

การประมวลผลข้อมูลจำนวนมากในเวลาอันรวดเร็วจำเป็นต้องใช้คอมพิวเตอร์ที่มีศักยภาพสูงและซอฟต์แวร์ระบบจัดการฐานข้อมูลที่มีประสิทธิภาพซึ่งล้วนแต่มีราคาแพง วิทยานิพนธ์นี้ได้เสนอแนวทางในการประยุกต์ใช้พีวีเอ็ม (Parallel Virtual Machine : PVM) กับการประมวลผลตัวดำเนินการจอยน์ (Join) เชิงขนานในระบบจัดการฐานข้อมูลโพสท์เกรสคิวแอล (PostgreSQL) โดยได้พัฒนาระบบต้นแบบบนระบบพีซีคลัสเตอร์ (PC Cluster System) เพื่อทดแทนคอมพิวเตอร์ประสิทธิภาพสูงและใช้ซอฟต์แวร์แบบเปิดเสรีรหัสโปรแกรมทดแทนซอฟต์แวร์เชิงพาณิชย์

การพัฒนาระบบต้นแบบได้นำพีวีเอ็มมาประยุกต์ใช้กับอัลกอริทึมหาแรลเลกไฮบริดแฮชจอยน์ (Parallel Hybrid Hash Join) โดยสถาปัตยกรรมของระบบต้นแบบถูกออกแบบให้สามารถทำการประมวลผลเชิงขนานระดับภายในตัวดำเนินการ (Intraoperator) จากผลการทดสอบระบบแสดงให้เห็นว่า ระบบที่พัฒนาขึ้นสามารถเพิ่มประสิทธิภาพในการประมวลผลจอยน์ อย่างไรก็ตามอัตราความเร็วที่เพิ่มขึ้นหยุดอยู่ที่ระดับหนึ่ง คือที่คอมพิวเตอร์จำนวน 13 เครื่อง เนื่องจากข้อจำกัดในการกระจายข้อมูล

Abstract

The very large-scale data can be processed faster by using high performance computers and efficient database management system (DBMS) that are expensive. This thesis purposes the implementation of parallel join operation in PostgreSQL DBMS using parallel virtual machine (PVM) in order to build a cheaper PC cluster system replacing the expensive high performance computers and employ the open source software instead of commercial software.

The prototype is designed and developed by applying PVM in Parallel Hybrid Hash Join algorithm which is the efficient parallel join algorithm. The prototype's architecture is designed by using the concept of intraoperator parallelism. The speedup and scaleup experiment of the developed system are tested. The result shows that the performance of Join operation is increased during adding up the computers into the system. However, the speedup curve is saturated when the number of computers added into the system are 13 due to the limitation of data distribution.