

ระบบจัดเก็บข้อมูลขนาดใหญ่ประสิทธิภาพสูงโดยใช้ พีซีคลัสเตอร์

คืออะไร

ระบบนี้ประกอบด้วยเครื่องคอมพิวเตอร์ส่วนบุคคล 4 เครื่องเชื่อมต่อกันโดยตรง โดยใช้ Ethernet Card และมีระบบ File Server ที่จัดการข้อมูลขนาด 1 เทราไบต์ (สามารถขยายได้ถึง 6 เทราไบต์) โดยมีซอฟต์แวร์ Distributed Queuing System (DQS) เป็นตัวจัดการการกระจายงานไปยังเครื่องต่างๆ และใช้ Software RAID ในการจัดการหน่วยจัดเก็บข้อมูลขนาดใหญ่ โดยการเข้าถึง ฮาร์ดดิสก์หลายๆ ตัวในเวลาเดียวกันเพื่อเพิ่มประสิทธิภาพในการอ่านและเขียนข้อมูล ระบบนี้ใช้ระบบปฏิบัติการ Linux และซอฟต์แวร์ต่างๆ ที่ใช้เป็น Open Software คือจัดทำขึ้นเพื่อเผยแพร่เป็นสาธารณะ จึงไม่มีค่าใช้จ่ายในส่วนซอฟต์แวร์ โดยทางเนคเทค ยังได้พัฒนาซอฟต์แวร์เพิ่มเติมเพื่อใช้จัดการระบบคอมพิวเตอร์นี้ด้วย

ลักษณะเด่นของระบบ

1. เป็นระบบที่ออกแบบมาเพื่อจัดการกับงานในลักษณะที่อาศัยการทำงานของหน่วยประมวลผลสูง และมีการเรียกใช้ข้อมูลขนาดใหญ่ ซึ่งรองรับตามความต้องการได้เป็นอย่างดี
2. ประสิทธิภาพในการต่อระบบเครือข่ายแบบนี้จะเร็วกว่าแบบที่ใช้ Network Switch เนื่องจากการเป็นการต่อถึงกันโดยตรงซึ่งไม่มีการสูญเสียในตัว Network Switch
3. ระบบนี้สามารถขยายได้ง่ายในอนาคต เมื่อความต้องการเพิ่มขึ้น ทั้งในแง่ของข้อมูลและประสิทธิภาพของการประมวลผล รวมทั้งจำนวนเครื่องคอมพิวเตอร์ (node) ที่มาต่อรวมกันในระบบ
4. ไฟล์เซิร์ฟเวอร์มีขนาดใหญ่ สามารถเก็บข้อมูลได้จำนวนมาก (ขยายได้ถึง 6 TB) ซึ่งรองรับการใช้งานของผู้ใช้ได้ไม่ต่ำกว่า 2 ปี
5. ระบบนี้มี Graphics Terminal 2 ชุด ทำให้ผู้ใช้ที่ Console ภายในสถานที่ที่สามารถใช้งานได้พร้อมกัน 2 คนโดยไม่ทำให้ประสิทธิภาพลดลง ระบบที่เสนอนี้ต่อเข้าเครือข่ายอินเทอร์เน็ตได้ 2 ช่องทางพร้อมกัน ช่วยเพิ่มอัตราการรับส่งข้อมูลให้สูงขึ้นสำหรับผู้ใช้งานที่ติดต่อจากภายนอก
6. มีการใช้โปรแกรมจัดการไฟล์เซิร์ฟเวอร์ เช่น RAID ย่อมาจากคำว่า Redundant Array of (Independent) Inexpensive Disks เป็นวิธีการเก็บข้อมูลให้กระจายไปในดิสก์หลายๆ ตัวเพื่อช่วยเพิ่มประสิทธิภาพในการอ่านและเขียนข้อมูล หรือเพื่อช่วยเพิ่มความน่าเชื่อถือในการเก็บข้อมูล หรือทั้งสองอย่าง
7. มีการใช้โปรแกรมจัดการการทำงานของโปรแกรม เช่น DQS (Distributed Queuing System) และ NAT (Network Address Translation) จะช่วยสร้างสมดุลของการทำงานของทุกเครื่องคอมพิวเตอร์ในระบบนี้ให้มีตัวใดตัวหนึ่งรับภาระมากเกินไป ตัวอย่างของโครงการที่ได้นำระบบนี้ไปใช้
8. มีการนำเอาโปรแกรมจัดการการเชื่อมต่อระบบไฟล์ เช่น Samba มาช่วยในการแลกเปลี่ยนข้อมูลระหว่างระบบปฏิบัติการ Linux และ Microsoft Windows

การนำไปใช้

ระบบนี้สามารถนำไปประยุกต์ใช้งาน เพื่อทดแทนระบบซูเปอร์คอมพิวเตอร์ หรือทำเป็นคลังข้อมูลขนาดใหญ่ของฐานความรู้ได้ อาทิเช่น

- ใช้ในการคำนวณทางคณิตศาสตร์ขั้นสูง เพื่อสร้างแบบจำลองต่างๆ
- นำไปใช้เป็นคลังข้อมูลขนาดใหญ่ ซึ่งใช้เก็บข้อมูลปริมาณมาก และเชื่อมต่อเข้าระบบเครือข่ายคอมพิวเตอร์ ทำให้เครื่องคอมพิวเตอร์อื่นๆ ที่ใช้ระบบปฏิบัติการต่างๆ กัน ยังสามารถเรียกใช้งานข้อมูลดังกล่าวได้
- ใช้ในการประมวลผลเพื่อสร้างฐานความรู้และระบบสนับสนุนการตัดสินใจ เช่น ระบบเครือข่ายข้อมูลสารสนเทศทางภูมิศาสตร์ เป็นต้น
- นำมาใช้เป็น WWW server โดยมีความสามารถในการจัดการรับส่งข้อมูล และรองรับผู้ใช้หลายคนในเวลาพร้อมๆ กันได้

เกร็ดความรู้

เรื่อง หน่วยความจุของการเก็บข้อมูลคอมพิวเตอร์ from <http://www.whatis.com>

Byte

ในระบบคอมพิวเตอร์โดยทั่วไป คำว่าไบต์เป็นหน่วยนับของข้อมูลที่มีขนาด 8 หน่วยเลขฐานสอง ซึ่งใช้แทนอักษร ตัวเลข หรือสัญลักษณ์ จำนวน 1 คำในระบบคอมพิวเตอร์

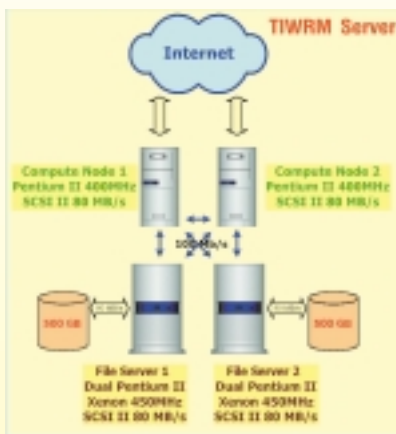
Megabyte

เป็นคำที่มีที่กบฏแล้วแต่การให้ความหมายโดยมีหลายที่มา เช่น

1. ตามความหมายของการจัดเก็บข้อมูลคอมพิวเตอร์ หมายถึง 2 ยกกำลัง 20 หรือเท่ากับ 1,048,576 ไบต์
2. ตามความหมายใน IBM Dictionary of Computing ใช้แทนข้อมูลปริมาณ 1,000,000 ไบต์
3. ตามความหมายของ Microsoft Press Computer Dictionary หมายถึง 1,000,000 ไบต์ หรือ 1,048,576 ไบต์ ก็ได้
4. ตามความหมายใน The New Hacker's Dictionary ของ Eric S. Raymond ใช้เป็น 1,048,576 ไบต์
5. บริษัท Iomega ได้เสนอผลิตภัณฑ์ชื่อ 100MB disk ซึ่งมีขนาดความจุเท่ากับ 100,431,872 ไบต์

Terabyte

หน่วยวัดของขนาดความจุของการเก็บข้อมูลในระบบคอมพิวเตอร์ ซึ่งเท่ากับ 2 ยกกำลัง 40 ไบต์ หรือเท่ากับ 1,099,511,627,836 ไบต์



ตัวอย่างระบบที่ใช้จริง



แสดงการเชื่อมต่อของระบบ