

80-81

超越计算环境的关键技术

李松树

(国防科技大学计算机系 长沙410073)

摘要 This paper gives the definition of metacomputing environment, superiorities, and in detail analyses its key technologies, at last, gives supposition for developing our country metacomputing environment.

关键词 Metacomputing, Metacomputing environment, Parallel program environment.

追求更高的劳动效率,促进了社会的不断向前发展,人类为此不仅使用了各种各样作为体力劳动增强器的机器,也越来越普遍地使用了作为脑力劳动辅佐的计算机,计算机已应用到人类社会的各个领域。同样是源于对高性能的追求,人们早在计算机发展的初期,就认识了并行处理的重要性,而且在计算机发展过程中也一直试图在系统的各个层次上开发并行性。近年来,高科技领域对计算机性能提出了越来越高的要求,而物理上的极限和工程实现上的问题使单机速度很难满足应用要求,人们迫切希望有一种能够解决任一单机解决不了的集超高速复杂计算、高精度、海量存储、视算于一体的并行计算环境——超越计算环境。

一、何谓超越计算环境

所谓超越计算环境是由多种不同结构的计算机通过网络互连方式组合成的一种复合计算环境,环境中可以包括标量计算机、工作站、向量巨型机、工作站群机、MPP巨型机、小巨型机等。这种环境可以提供一般单一体系统结构计算机不具有的集细粒度、中粒度和粗粒度计算,视算服务和海量存储于一体的异构计算环境,它具有的综合计算和处理能力是单机不可比拟的,因此在更大程度上满足了人们对高性能计算的要求。

二、超越计算环境所具有的优势

超越计算环境具有如下优势:

1. 灵活性。超越计算环境能方便地将不同体系结构的各种计算环境互连在一起,有机地形成一个用户所需的异构计算环境。

2. 适用多种计算的超高速高精度复杂计算和仿真。环境中的各种体系结构的节点计算机使得它既可以进行向量超高速计算,也可以进行标量超高速计算,同时各节点机还可以基于网络进行并行处理,通过粗粒度,中粒度,细粒度的并行,协同计算共同

解算高复杂度的计算问题。例如工作站具有很强的图形处理能力,能够满足交互类问题的需要,工作站群机可以在很大程度上满足大吞吐率问题的要求,而小巨型机, MPP 则用来解算处理能力类问题。

3. 海量数据存储和存储管理。海量的磁盘子系统 and 磁带子系统通过高速通道与各计算资源相连形成数据中心。

4. 视算。系统在具有高带宽 I/O 的大容量数据存储管理和向量巨型机, MPP 的后备支持下,通过工作站提供各种图形,声像及多媒体处理,也可以可视化程序并行执行情况。

5. 超高作业吞吐量。支持更多的用户上机,缩短响应时间。

6. 超高可靠性。即使环境中有一个节点或多个节点出现故障,系统仍能降级使用。

三、超越计算环境的关键技术

超越计算环境的综合优势要求构成环境的物质基础——单机性能最优,因而单机组合的超越计算环境才有可能最优。作为一种异构的多种计算机系统,互连网络,并行程序环境等是实现超越计算环境的关键技术。

1. 互连网络

互连网络是超越计算环境的重要组成部分,环境中的各计算机均由互连网络互相连接,互相通信,一个好的互连网络应同时具有:①稳定的高传输速率(100~800Mb/s);②低延迟(收发数据的软件开销在毫秒级);③良好的可延伸性;④支持广播通信;⑤易于从网络故障中恢复重新运行。

其中,两个主要因素是网络延迟和传输速率。网络延迟是指系统发送/接收一个长度消息时系统开销的测度,传输速率是指系统在单位时间内发送/接收的数据的数量,它是对系统数据流量的测度,互连网络可以有多种形式,适用于不同的应用。

(1)商用局域网,如 Ethernet, FDDI, 特点是互连

李松树 高级工程师,主要从事大规模并行计算机的操作系统、编译系统的研究。

简单,投资少,但网络延迟是MPP的30倍,传输速率相差更多。这种通讯机制严重限制了超越计算环境的应用范围,因而只适用于对大吞吐量要求的应用领域。

(2)高速互连接口,如convex公司的SMI,其通讯带宽达14MB/S,网络延迟小于0.5ms,其性能出色,适用于高速计算环境。

(3)基于ATM的通信网络,其网络延迟和传输速率性能比SMI更高,适应于超越计算环境。

不论是采用专门的互连网络,还是利用标准的通信网络(SMI,ATM),关键是要解决接口技术问题、研制精简而高效的网络协议,节点机之间的通信不应进入操作系统,通讯与计算尽可能重迭。

2. 高速缓存及存储管理

超越计算的依据是程序执行中空间与时间的局部性。由于SRAM成本太高,而Cache速度随其容量增大而下降,高速缓冲的容量受限制,Cache一致性的硬件支持是实现具有统一地址空间的分布式共享存储的基础。必须要有专门的存储管理部件来保证各个处理机共享统一的地址空间。

3. 高I/O能力

磁盘速度的提高,严重落后于CPU速度的提高,I/O速度已成为并行机明显的瓶颈,严重影响了并行系统的性能。应采用如下技术提高I/O能力:(1)大量的小磁盘组成冗余磁盘阵列;(2)数据压缩技术。

4. 并行编译系统

并行编译系统的发展落后于体系结构的发展,目前的并行编译系统不论是对共享变量、消息传递还是数据并行程序设计模式的支持都不理想,今后我们应下大力气研究并行优化编译器,对其并行性识别,数据相关,数据分配,并行划分,程序常规优化与重构,并行优化等技术进行重点研究,力争使并行编译系统水平上一个新台阶。

5. 操作系统的标准化

构成超越计算环境的各计算机,一般要求相同的系统平台,以实现各计算机之间的透明互操作,选用unix作为操作系统的标准平台是合适的,由于unix采用由微内核技术实现的Mach作核心,支持多线程技术和消息传递机制,这种IPC机制可以有效地支持松耦合多机系统,支持超越计算环境的细粒度并行。

6. 系统管理工具

超越计算环境应设有一套有效的系统管理工具,使得系统管理人员可以把整个系统看成一个或多个逻辑设备,系统管理员使用一套命令集就可以对整个系统进行有效的管理。

7. 并行程序环境

并行程序环境作为减轻用户在并行方面的困

难,协调开发各阶段活动的有效手段正在受到广泛的关注。它是实现超越计算的重要基础。并行程序环境包括并行程序设计环境和并行程序运行环境,其作用是为用户在使用超越计算环境解算应用问题时提供描述程序并行性的手段,对程序运行时的各处理机间的通信和负载平衡等给予支持。并行程序环境由如下组成:(1)并行程序设计语言;(2)消息传递库;(3)并行编程模型,主要有共享变量、消息传递和数据并行;(4)并行工具集,主要有:自动并行化工具、程序抽象表示工具、源级程序调试工具、程序性能监视、分析、调整工具、程序执行示踪工具、作业/任务调度及自动负载平衡工具。

并行程序环境是用户最关心的人机界面,一直是并行处理领域中研究的热点。它是提高计算能力,降低计算成本的最终出路,必将计算机跨世纪研究发展的主流方向。近年来,支持超越计算环境的软件包不断面世,但性能最好,影响最大的是PVM和EXPRESS,它们都是基于网络的并行程序环境,支持C语言和FORTRAN语言,它们以独立于具体机器结构的方式提供软件生存周期所需要的所有工具,包括分析、设计、编码、调试、性能评估、维护等工具,并且为大多数工具提供了图形界面。EXPRESS性能比PVM高,但它是一种商品化软件,而PVM则可免费获得。

四、新的机遇和挑战

超越计算环境是实现高性能计算的一种有效途径,它所具有的多种计算资源的综合优势和高速计算、概算以及海量存储与管理能力使得它具有广阔的应用前景。目前,在国外,超越计算很热,这对于我国的计算机界是一次机遇,也是一次挑战。我国的高性能计算落后于外国,我们能不能在高性能计算机市场上占一席之地?我们要不要在高性能计算机市场上占一席之地?我们怎样在高性能计算机市场上占一席之地?笔者认为,中国的计算机工作者一定有所作为。对于第一问,我们知道,并行机硬件最基本的部件是微处理器芯片和DRAM存储芯片,在开放条件下,这些器件从国外的市场上可以买到。并行机最关键的技术是并行软件,从国外来看,还没有那一家能够垄断并行软件市场,许多厂商也才开始研究,事实上,中国人的脑袋并不比外国人笨。只要我国国内同行相互协作,借鉴国外的先进技术,攻击互连网络和并行程序环境等关键技术,我们一定能在高性能计算领域占有一席之地。对于第二问,回答是肯定的。对于第三问,我们应该以理论和实践相结合为指导思想,采用“拿来”、“继承”和“创新”相结合的方法,使我国的高性能计算起点高,进展快,收效好。这样才能缩短我国与国外的差距,使我国的并行处理水平跃上一个新台阶。