

第1章 预备知识

本章主要介绍学习并行计算所需要了解的一些基础知识。首先,给出并行计算的一个简单定义,并讨论并行计算的主要研究目标和内容,以及推动并行计算发展的主要动力。其次,简要介绍并行计算赖以存在的硬件平台——并行计算机的发展史。然后,简要介绍当前并行计算机的体系结构,以及运行在这些并行计算机上的操作系统和并行编程环境,使读者对并行计算机系统有一个大致的了解,为深入学习并行编程奠定基础。最后,简要讨论并行算法。

1.1 并行计算的主要研究目标和内容

并行计算是伴随并行计算机的出现,在近30年来迅速发展的一门交叉学科,涵盖的内容非常广泛。参考文献[1]较为全面地综述了并行计算在各个方面的最新进展,内容包括并行计算机体系结构、编译系统、并行算法、并行编程、并行软件技术、并行性能优化与评价、并行应用等。因此,很难全面并精确地给出并行计算的定义。但是,从交叉学科的角度,并行计算可以定位为连接并行计算机系统和实际应用问题之间的桥梁。它辅助科学、工程及商业应用的领域专家,为在并行计算机上求解本领域问题提供具有共性的关键支撑。本节基于这点认识,首先给出并行计算一个粗浅的定义,然后讨论并行计算的主要研究目标和内容,以及推动并行计算发展的主要动力。

1.1.1 什么是并行计算

并行计算 (parallel computing) 是指在并行计算机上, 将一个应用分解成多个子任务, 分配给不同的处理器, 各个处理器之间相互协同, 并行地执行子任务, 从而达到加快求解速度, 或者提高求解应用问题规模的目的。

因此, 为了成功开展并行计算, 必须具备 3 个基本条件:

(1) 并行计算机。并行计算机至少包含两台或两台以上处理机, 这些处理机通过互连网络相互连接, 相互通信。

(2) 应用问题必须具有并行度。也就是说, 应用可以分解为多个子任务, 这些子任务可以并行地执行。将一个应用分解为多个子任务的过程, 称为并行算法的设计。

(3) 并行编程。在并行计算机提供的并行编程环境上, 具体实现并行算法, 编制并行程序并运行该程序, 从而达到并行求解应用问题的目的。

1.1.2 并行计算的主要研究目标和内容

对于具体的应用问题, 采用并行计算技术的主要目的在于两个方面:

(1) 加快求解问题的速度。例如, 给定某应用, 在单处理器上, 串行执行需要 2 个星期 (14 天), 这个速度对一般的应用而言, 是无法忍受的。于是, 可以借助并行计算, 使用 100 台处理器, 加速 50 倍, 将执行时间缩短为 6.72 小时。

(2) 提高求解问题的规模。例如, 在单处理器上, 受内存资源 2GB 的限制, 只能计算 10 万个网格, 但是, 当前数值模拟要求计算 1000 万个网格。于是, 也可以借助并行计算, 使用 100 个处理器, 将问题求解规模线性地扩大 100 倍。

并行计算之所以必需, 主要在于当前的单处理器性能不可能满足大规模科学与工程计算及商业应用的需求, 并行计算是目前

惟一能满足实际大规模计算需求的支撑技术。例如,即使是当前较为先进的微处理器 (Itanium2 1.5GHz、POWER4 1.5GHz 等),其峰值性能也仅为 60 亿次/s。近 2 年内,微处理器的峰值性能也不会超过 100 亿次/s。并行计算之所以可行,主要在于,并发性是物质世界的一种普遍属性,具有实际应用背景的计算问题在许多情况下都可以分解为能并行计算的多个子任务。

综上所述,并行计算的主要目标在于,在并行计算机上解决一批具有重大挑战性计算任务的科学、工程及商业计算问题,满足不断增长的应用问题对速度和内存资源的需求。

由并行计算的 3 个必备条件可知,并行计算的主要研究内容大致可分为 4 个方面:

(1) 并行计算机的高性能特征抽取。主要任务在于,充分理解和抽取当前并行计算机体系结构的高性能特征,提出实用的并行计算模型和并行性能评价方法,指导并行算法的设计和并行程序的实现。

(2) 并行算法设计与分析。针对应用领域专家求解各类应用问题的离散计算方法,设计高效率的并行算法,将应用问题分解为可并行计算的多个子任务,并具体分析这些算法的可行性和效果。

(3) 并行实现技术,主要包含并行程序设计和并行性能优化。基于并行计算机提供的并行编程环境,例如消息传递平台 MPI^[20]或者共享存储平台 OpenMP^[8],具体实现并行算法,研制求解应用问题的并行程序。同时,结合并行计算机的高性能特征和实际应用的特点,不断优化并行程序的性能。

(4) 并行应用。这是并行计算研究的最终目的。通过验证和确认并行程序的正确性和效率,进一步将程序发展为并行应用软件,应用于求解实际问题。同时,结合实际应用出现的各种问题,不断地改进并行算法和并行程序。一个没有经过实际应用验证和确认的算法和程序,不能说是一个好的算法和程序。

以上 4 个方面相互耦合,缺一不可。(1) 是开展并行计算研究的基础,(2)、(3) 是并行计算研究的核心,也是在并行计算机上高效求解应用问题的基本保证,(4) 是并行计算研究的目的,也是验证和确认并行计算研究成果的最有效的途径。(4) 将为 (2)、(3) 的研究提供取之不尽的源泉。

需要说明的是,并行计算不同于分布式计算 (distributed computing)。后者主要是指,通过网络相互连接的两个以上的处理机相互协调、各自执行相互依赖的不同应用,从而达到协调资源访问,提高资源使用效率的目的。但是,它无法达到并行计算所倡导的加快求解同一个应用的速度,或者提高求解同一个应用的问题规模的目的。对于一些复杂应用系统,分布式计算和并行计算通常相互配合,既要通过分布式计算协调不同应用之间的关系,又要通过并行计算提高求解单个应用的能力。

下面给出并行计算和分布式计算的几个例子,以示区别。

例 1.1 (并行计算) N 个数被分布存储在 P 台处理器, P 台处理器并行执行 N 个数的累加和。首先,各个处理器累加它们各自拥有的局部数据,得到部分和;然后, P 台处理器执行全局通信操作,累加所有部分和,得到全局累加和。

例 1.2 (并行计算) 给定二维规则区域上的 Dirichlet 问题 $-\Delta u = f$, 采用标准 5 点有限差分格式离散。平均分配 $N = N_x \times N_y$ 个网格单元给 $P = P_x \times P_y$ 台处理机。所有处理机并行计算,执行 Jacobi 迭代,求解 Dirichlet 问题,从而达到缩短求解问题的时间,或者扩大网格规模 N 的目的。

例 1.3 (分布式计算) 观众点播、远程驾驭式可视化、电视会议等。

1.1.3 推动并行计算发展的主要动力

大规模科学与工程计算应用对并行计算的需求是推动并行计算快速发展的主要动力。长期以来,它们对并行计算的需求是无

止境的。例如,全球气象预报中期天气预报模式要求在 24 小时内完成 48 小时天气预测数值模拟,此时,至少需要计算 635 万个网格点,内存需求大于 1TB,计算性能要求高达 25 万亿次/s。又如,美国在 1996 年开始实施 ASCI 计划,要求分 4 个阶段,逐步实现万亿次、10 万亿次、30 万亿次和 100 万亿次的大规模并行数值模拟,实现全三维、全物理过程、高分辨率的核武器数值模拟。除此之外,在天体物理、流体力学、密码破译、海洋大气环境、石油勘探、地震数据处理、生物信息处理、新药研制、湍流直接数值模拟、燃料燃烧、工业制造、图像处理等领域,以及大量的基础理论研究领域,存在计算挑战性问题,均需要并行计算的技术支持。

近 10 年来,美国在大规模科学与工程计算应用领域,启动了 3 次重大计划,极大地推动了并行计算的发展。通过这 3 次计划,美国无论在并行计算机的研制方面,还是在并行应用程序的开发方面,均取得了国际领先的绝对优势,极大地推动了并行计算在科学与工程计算各个领域的应用,全面加速了科学技术的发展。

第一次是在 1983 年,配合星球大战战略防御计划而开展的战略计算机计划 (SCP),研制了每秒 10 亿次的 CRAY 并行计算机。

第二次是在 1993 年,由美国科学、工程、技术联邦协调委员会向国会倡议的“重大挑战性项目:高性能计算与通信 (HPCC) 计划”,目的是研制能够提供 3T 性能目标 (1TFLOPS 计算能力、1TB 内存容量和 1TB/s 的 I/O 带宽, $1T=10^{12}$) 的高性能并行计算机,解决科学与工程计算中的重大挑战性课题,保持其在高性能计算和计算机通信领域中的世界领先地位,利用高性能计算机与网络技术刺激生产,以提高国民经济、国家安全、教育和整体环境的竞争力。该计划由高性能计算机系统 (HPCS)、先进算法与软件 (ASTA)、国家科研与教育网 (NREN)、基本研究与人类资源 (BRHR) 4 部分组成,其目标是:

- (1) 对一大批重要应用问题, 计算性能要达到每秒万亿次

运算;

- (2) 发展相关的系统软件, 对一大批问题改进算法;
- (3) 国家研究网能力要达到每秒 10 亿位;
- (4) 充分保证计算机科学与工程领域的科研人员的需求。

HPCC 计划提出的背景是一大批巨大挑战性问题需要解决, 其中包括: 天气与气候预报, 分子、原子与核结构, 大气污染, 燃料与燃烧, 生物学中的大分子结构, 新型材料特性, 国家安全等有关问题。而近期内要解决的问题包含: 磁记录技术、新药研制、高速城市交通、催化剂设计、燃料燃烧原理、海洋模型模拟、臭氧层空洞、数字解剖、空气污染、蛋白质结构设计、金星图像分析和密码破译技术。世界上第一台峰值速度超过 1TFLOPS 的高性能计算机由 Intel 公司于 1996 年 12 月成功研制。

日本相继提出了真实世界计算计划, 欧洲提出了万亿次机计划, 我国 863 高科技计划也将并行计算机的研制列为关键攻关技术。

第三次是在 1996 年, 美国能源部联合美国三大核武器实验室 (Los Alamos 国家实验室、Lawrence Livermore 国家实验室和 Sandia 国家实验室) 共同提出了“加速战略计算创新 (ASCI) 计划”, 提出通过数值模拟评估核武器的性能、安全性、可靠性、更新等, 要求数值模拟达到高分辨率、高逼真度、三维、全物理、全系统的规模和能力。为此, 三大实验室分别向美国三大公司 (Intel、IBM 和 SGI) 预订了峰值速度超过 1TFLOPS 的并行计算机, 计划分 4 个阶段, 分别实现 1 万亿次、10 万亿次、30 万亿次和 100 万亿次的高性能并行计算机。目前, 以 TOP 500^[9] 中排名第一的峰值性能达 185 万亿次的 IBM Blue Gene/L 为标志, 4 个阶段的并行计算机研制已经初步实现。目前, 整个 ASCI 计划发展到第 2 个阶段, 改名为“先进模拟计算计划 (ASC)”。

对应于美国的 ASCI 计划, 日本和欧洲也提出了相应对策。尤

其是日本,以 2002 年研制的 Earth Simulator^[9] 及其在大气海洋环流等大规模科学与工程计算中的高效率应用为标志,将并行计算推向了一个新的高度。

除了大规模科学与工程计算应用外,微电子技术与大规模集成电路 VLSI 的发展是推进并行计算发展的另一个主要动力。当前,Moore 定律仍在延续,计算机微处理器的速度每 3 年翻两番,内存容量每两年翻 3~4 倍;互联网络技术飞速发展,专用并行计算机网络的延迟可低于 $1\mu\text{s}$,带宽可达 6.4GB/s ,商用并行计算机网络延迟可低于 $6\mu\text{s}$,带宽可达 1.25GB/s 。这些基本构件性能的改进,无疑会改变并行计算机的体系结构、操作系统和编译系统、并行编程环境,从而对并行算法和并行编程模式带来新的挑战。当前,如何快速地开发并行应用程序,高效率地发挥当前并行计算机的峰值性能,已经成为当前并行计算研究面临的一个挑战性问题。

1.2 并行计算机发展历史

并行计算机从 20 世纪 70 年代的开始,到 20 世纪 80 年代的蓬勃发展和百家争鸣,20 世纪 90 年代体系结构框架趋于统一,近 5 年来机群技术的快速发展,并行计算机技术日趋成熟。本节以时间为线索,简介并行计算机发展的推动力和各个阶段,以及各类并行计算机的典型代表和它们的主要特征。

1.2.1 应用需求的推动作用

市场需求一直是推动并行计算机发展的主要动力,大量实际应用部门,例如数值天气预报、核武器、石油勘探、地震数据处理、飞行器数值模拟和大型事务处理、生物信息处理等,都需要每秒执行万亿次、数十万亿次乃至数百万亿次浮点运算的计算机。正如引言中所指出的,基于这些应用问题本身内部存在的并行性和单机性能的限制,并行计算是满足它们需求的惟一可行途径。

高性能并行计算机的研制与应用水平一直是以美国和日本为首的各个发达国家共同追逐的目标,是衡量一个国家科技、经济和国防综合实力的重要标志。在以明确应用需求为背景的 HPCC 计划和 ASCI 计划的推动下,美国在这方面取得了绝对的领先地位。

1.2.2 20 世纪 70 年代

1972 年,世界上诞生了第一台并行计算机 ILLIAC IV,它含 32 个处理单元,环形拓扑连接,每台处理机拥有局部内存,为 SIMD 类型机器。对大量流体力学程序,ILLIAC IV 获得了 2~6 倍于当时性能最高的 CDC 7600 机器的速度。

20 世纪 70 年代诞生的并行计算机还有阵列机 ICLDAP、Good-year MPP,以及向量机 CRAY-1、STAR-100 等,它们都属于 SIMD 类型,其中向量机 CRAY-1 获得了很好的向量计算效果。20 世纪 70 年代的并行计算机引起了人们的极大兴趣,吸引了大量的专家学者从事并行计算机研制和并行程序的设计,为 20 世纪 80 年代并行计算机的蓬勃发展奠定了坚实的基础。

1.2.3 20 世纪 80 年代早期

20 世纪 80 年代早期,以 MIMD 并行计算机的研制为主。首先诞生的是 Denelcor HEP,含 16 台处理机,共享存储,能同时支持细粒度和大粒度并行,并且被应用到实际计算中,使许多人学会了并行计算。其次,诞生了共享存储向量多处理机 CRAY X-MP/22(2 个向量机结点)、IBM 3090(6 个向量机结点),取得了很好的实际并行计算性能。同时,以超立方体结构连接的分布式存储 MIMD 结构原型机开始出现。

1.2.4 20 世纪 80 年代中期

20 世纪 80 年代中期,共享存储多处理机系统得到了稳定发展。两个成功的机器为 Sequent (20 个结点)、Encore(16~32 个结