

网格与商务智能

WANGGE YU SHANGWU ZHINENG

章剑林 著



上海交通大学出版社

网格与商务智能

章剑林 著

上海交通大学出版社

内 容 提 要

网格计算是建立在互联网技术基础上,针对复杂科学计算的一种新型计算模式,它与经济、商业需求和发展密切相关。本书从信息技术、管理科学等角度研究了网格技术在商务领域多方面的应用,在阐述网格技术、网格体系结构、网格商务及其与电子商务的关系、网格商务中的市场营销、网格商务中的供应链管理、网格商务中的知识管理、商务智能技术等内容基础上,重点研究了商务领域应用中网格计算模式与现代商务智能技术相结合问题,实现了网格计算与智能技术在电子商务领域中的应用,为促进企业网格和 BI 项目的实现提供一个良好的理论和技术基础。本书在撰写过程中尽可能从基本原理分析入手,再提出解决问题的方法,并以具体应用实例加以说明。

本书可供网格技术、商务智能技术及其商务领域应用的专业人员和相关管理人员参考,也可供高等学校有关专业的学生学习之用。

图书在版编目(CIP)数据

网格与商务智能/章剑林著. —上海:上海交通大学出版社,2008
ISBN978-7-313-05369-5

I. 网... II. 章... III. 网格—计算技术—应用—商业管理 IV. F715

中国版本图书馆 CIP 数据核字(2008)第 136860 号

网格与商务智能

章剑林 著

上海交通大学出版社出版发行

(上海市番禺路 951 号 邮政编码 200030)

电话:64071208 出版人:韩建民

上海交大印务有限公司 印刷 全国新华书店经销

开本:787mm×1092mm 1/16 印张:11.5 字数:279 千字

2008 年 9 月第 1 版 2008 年 9 月第 1 次印刷

印数:1~2050

ISBN978-7-313-05369-5/F·780 定价:25.00 元

版权所有 侵权必究

序

自 1998 年美国 Argonne 国家实验室和芝加哥大学教授 Ian Foster 和 Carl Kesselman 所著《网格——崭新计算基础设施的蓝图》一书以来,已近 10 年时间。计算网格作为一种计算机硬件和软件的公共设施,向用户提供可靠、一致、广泛和廉价的高端计算能力,已经越来越成为现代社会、经济和生活的重要部分。

电子商务是 21 世纪全球经济发展的一道崭新的风景。随着互联网络技术的不断普及和经济全球化进程的不断加快,人们之间的时空距离大大缩短,信息交流与沟通展现出前所未有的频繁与顺畅,这不仅对人们的生活方式产生了深远的影响,也对企业的经营方式、商业模式产生了巨大冲击。被称为第三代互联网的网格技术的发展,为新世纪电子商务模式的创新与变革提供了巨大的契机和动力。网格计算技术为建立在互联网络基础平台上的电子商务模式的创新在技术上有了保证,同时,网格的丰富思想对电子商务模式提供了非常重要的经营与管理的思想源泉。

目前涉及有关网格计算内容的研究多是网格技术理论及其科研领域,应用领域也多是与国外科研机构项目合作密切的科学研究和生物技术等项目,真正深入涉及商业领域的计算网格研究尚不多见;有关商务智能方面的论著也大多是商务智能技术中所涉及的几种经典技术,外加一些软件开发新技术平台,缺少新的商务智能模型的开拓。

本书从网格计算与商务智能的结合点出发,研究涉及网格计算与商务智能相结合的部分,目前这方面内容的论著较少,而这个命题是目前商务领域应用发展过程中必须面临的问题。

本书的出版对丰富网格计算和商务智能领域理论研究,促进相关学科发展具有十分积极的意义。一方面,有助于该领域多学科的融合和发展;研究内容涉及计算机科学、信息科学、经济学和管理科学等多个学科,从经济学和管理科学的视角,运用计算机科学和信息科学相关理论和技术手段对现代商务领域应用进行研究。有关网格计算和商务智能的研究,是当今两大学科的发展前沿和研究热点,所以在研究中还结合了经济学和管理科学知识的应用。另一方面,开拓了网格商务智能技术新领域:在探讨该领域多学科理论的基础上,重点开展了对网格商务智能技术的研究,结合前人在网格计算和商务智能方面的研究成果,面向现代商务发展需求,找到研究结合点,开拓了网格商务智能技术的研究新领域,有助于信息技术在现代商务领域的推广应用。

本书对培养该领域的研究和管理人才具有积极推动作用。我国网格计算和商务智能领域研究起步较晚,该领域专业人才缺少,这在该领域的研究和教学中就可以感受到。本书贯穿由浅入深的思路,从网格计算和商务智能的基本概念开始,使其逐步深入到经济学、管理科学等多学科相关领域中,直到最后才提出网格商务智能技术与其相关项目的设计实施,这必将为相关专业人士参考和学生的学习提供便利。

陈德人

2008 年 8 月

前言

网格计算是一项针对复杂科学计算的新型计算模式,它将互联网的应用和发展推向一个崭新的阶段。在费用没有明显增长的前提下,网格计算系统可以帮助用户完成以前难以承担的任务,有效地解决了为 IT 基础架构所支付的巨大费用这个一直困扰企业发展的难题,并在资源共享和应用集成方面表现出较强的能力。其精确而快速的计算能力、数据存储和管理能力、支持决策分析能力、提高资源利用率的能力、提高员工工作效率、提高与其他组织协同解决问题的能力、一站式服务的能力,让人们为之惊叹。网格以建立高效灵活、低成本的组织架构为己任,成为解决企业信息技术应用的有效工具。

同时,网格技术与企业的经营管理活动相结合,催生了企业经营管理的新模式——网格计算环境下的商务模式。网格商务技术的运用大大促进了电子商务的进一步发展,运行电子商务的企业不仅可共享信息资源,而且也可共享空间网格计算资源,其运行模式、管理理念、应用环境也因此随之改变。

目前,国际上许多国家的众多著名科研机构开始了网格商务应用的研究,网格已经成为应用研究的热点,很多企业也纷纷开始实施商务网格项目,为商业网格应用赢得先机。我国对网格技术的研究虽然已有一段时间,但涉及在教育、科研领域以及有关商务领域的具体应用尚不多见。

商务智能作为现代企业的常用技术手段正逐步进入到各个领域,许多企业开始利用商务智能技术来解决典型的商业问题,诸如直接营销、面向细分市场客户群体划分、用户背景分析、交叉销售等市场分析行为,以及客户流失性分析和客户信用评价等,也逐步渗入到企业资源规划、客户关系管理、供应链管理、财务和人力资源管理等各个方面。在日趋紧张的商业经济环境下,企业如何生存发展,关键在于它是否能够为各种不同用户需求作出快速的反应以及正确的决策,并提供优秀的服务和产品。商务智能的实质是从大量数据中有效地提出信息,从信息中及时发现知识,在正确的时间将正确的结论交给合适的人们,从而支持决策过程。因此,在商业智能系统中需要大量有用的、精确的数据及适当的分析工具,来更好地解决商务中的相关问题。目前,人们对商务智能的重要性已经达成共识,有关商务智能的研究也十分活跃。

随着现代商务理论和应用研究的进一步深化及商务应用实践活动的快速发展,网格环境下的智能商务模式和运行机制必将有更大的创新,为现代商务成长提供理论和方法指导。目前,网格计算和商务智能还只是处在初步应用阶段,很多技术还在实验阶段,网格环境下的智能商务模式的形成还需要经过一个不断成熟的过程。

本书内容是作者近年来这方面工作的总结。书中研究了网格技术在商务领域的应用,在此基础上,重点研究了商务领域应用中网格计算模式与现代商务智能技术相结合的问题,并就网格技术与智能技术在电子商务领域中的应用,构建了网格商务智能系统平台原型,开发了企业信息网格和商务智能相结合的项目,实现了企业信息网格和 BI 项目的统一,为规范该领域项目管理,促进企业信息网格、BI 项目过程改进和质量管理,提供了一个良好的理论和技术基础。

本书的撰写得到了很多朋友和同事的关心和帮助。在此,特别要感谢浙江大学的陈卫东博士,正是有了他的无私帮助和对本书的一直关注,本书才得以顺利完成;同时还要感谢浙江大学的徐从富博士以及王昌建、田美霞、傅俊和赵瑞芬等各位同事。本书第7章部分内容涉及作者与陈卫东博士在网格和商务智能领域合作的相关内容。

感谢我的妻子和孩子,他们为我付出了很多。本书中如有不当之处,敬请各位专家和同行批评指正。

章剑林

2008年6月

目 录

第 1 章 网格基础	1
1.1 网格的概念	1
1.2 网格的产生与发展	4
1.3 网格的特点	9
1.4 网格的分类	10
1.5 网格与商务	13
第 2 章 网格的基本结构	16
2.1 网格的体系结构	16
2.2 五层沙漏体系结构	19
2.3 OGSA 体系结构	22
2.4 Web 服务资源框架	27
2.5 网格的基本应用	35
第 3 章 网格技术	42
3.1 网格的资源管理	42
3.2 网格的信息管理	50
3.3 网格的数据管理	54
3.4 网格的安全技术	58
3.5 Web Services	65
第 4 章 网格商务	71
4.1 网格商务的概念	71
4.2 网格商务与电子商务	73
4.3 网格商务对经济领域的影响及其自身的特点	75
4.4 网格商务中的市场营销	79
4.5 网格商务中的供应链管理	87
4.6 网格商务中的知识管理	96
第 5 章 商务智能	102
5.1 商务智能的基本知识	102
5.2 商务智能的发展历史	103
5.3 商务智能的体系结构	106

5.4 商务智能系统与其他系统	107
5.5 网格商务智能	109
第6章 商务智能技术	110
6.1 商务智能管理技术	110
6.2 数据 ETL 处理	114
6.3 数据仓库	119
6.4 联机分析技术	127
6.5 数据挖掘	129
6.6 联机分析挖掘	133
6.7 可视化技术	136
第7章 基于网格的商务智能系统开发	139
7.1 基于网格的商务智能服务平台	139
7.2 系统需求分析及开发	140
7.3 功能特性	143
7.4 系统实现的技术分析	148
7.5 技术难点	159
7.6 系统的 UML 图	160
7.7 系统的数据库设计	165
参考文献	174

第 1 章 网络基础

1.1 网络的概念

20 世纪最伟大的发明之一就是网络,它改变了人们的生存方式。多年前人们已经通过网络将各种传感器、计算机和相关数据信息连在一起,近年来我们又看到了日益广泛的网格应用。随着网格技术的应用,原有的交互模式和酬劳模式得到改变,同时我们也体会到网格协同工作所带来的非线性增长的巨大好处。协同可以使人类的劳动成果得到奇迹般的放大,而网格的成功将最终促成和依赖协同这样的放大器。

1.1.1 什么是网格

随着 Internet 网络技术的发展,一方面由于网络带宽的不断加大和计算机处理数据性能的不断增强,很多计算机资源只用于简单的文字和数据处理,其利用效率远没有得到发挥;另一方面,由于互联网上的信息资源每天都在不断增长,很多信息都仅仅是数量上的增长,据有关专业人士估计目前每年增长的信息达 2×10^{18} B,而为公众所用的信息只占总量的 0.000 15%,即便是 Google 这样功能强大的搜索引擎也只能搜查 1.3×10^8 B 的内容。如何充分利用这些资源是目前为大家所关注的问题。网格(Grid)技术就如何更好地利用现有各种计算机软硬件资源和网络信息资源提出了一个很好的解决思路。

网格通过通信手段将有关资源无缝集成为一个有机整体,可向用户提供一个基于国际互联网的新型计算平台。在这个平台上对来自不同客户的请求和提供资源的能力之间进行合理的匹配,为用户的请求选择合适的自愿服务,实现广域范围的资源共享。网格也可以看作是一架性能巨大的超级计算机,提供了包括计算资源、存储资源、数据资源、信息资源、知识资源、专家资源、领域资源和设备资源等资源在内的全面共享和互操作服务。

伊安·福斯特(Ian Foster),美国阿岗国家实验室的资深科学家,网格领域最有影响的人之一,在《网格:一种未来的计算基础设施蓝图》一文中这样描述网格:“网格是构筑在互联网上的一组新兴技术,它将高速互联网、计算机、大型数据库、传感器、远程设备等融为一体,为科技人员和普通百姓提供更多的资源、功能和服务。互联网主要为人们提供电子邮件、网页浏览等通信功能,而网格则提供更多更强的功能,它能让人们共享计算资源、存储资源和其他资源。”

网格的英文 Grid 含义源于电力网的英文 Power Grid 的思想。在电力网(见图 1-1)中,当人们插上电源插座,打开开关就能得到所需要的电力,完成相应的操作,而根本不用考虑也不可能区分所使用的电力是来自核电还是水电或火电。因为在电力网中所有电能通过各种输送变电设备,已经合并成一个能提供电力能源的综合网络。

网格(见图 1-2)借助了电力网的这种思想,将网络上所有资源综合集成起来,向所有用户

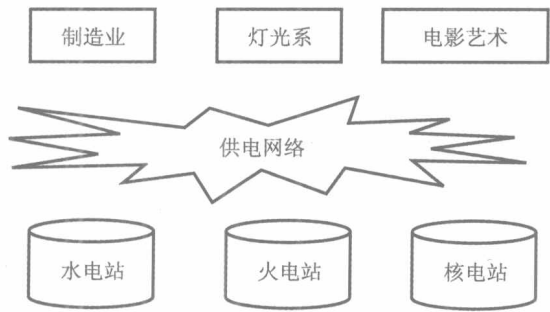


图 1-1 电力网

提供所需的服务。一方面，网格平台将各种包括超级计算机、海量存储器、数据库、网络和各种设施设备综合集成起来，以统一的标准描述其功能和访问接口，并按照注册中心登记后对外进行信息服务。任何一台计算机通过向各注册中心的注册登记，自动检索所需的服务。按照网格协议，计算机之间的资源集成起来，就能相互传送和阅读各种信息格式和相关内容，自动完成信息的融合，形成一个满足应用需要的工作流程，向用户提供前所未有的超级

服务。同时，网格还通过整套的资源管理、错误检测和安全保证机制，确保系统安全高效地运行。另一方面，网格平台还提供各种支持不同类型应用的常用工具，如协同工作工具、数据管理工具和分布式仿真工具等。这些工具将大大方便基于网格的应用开发。

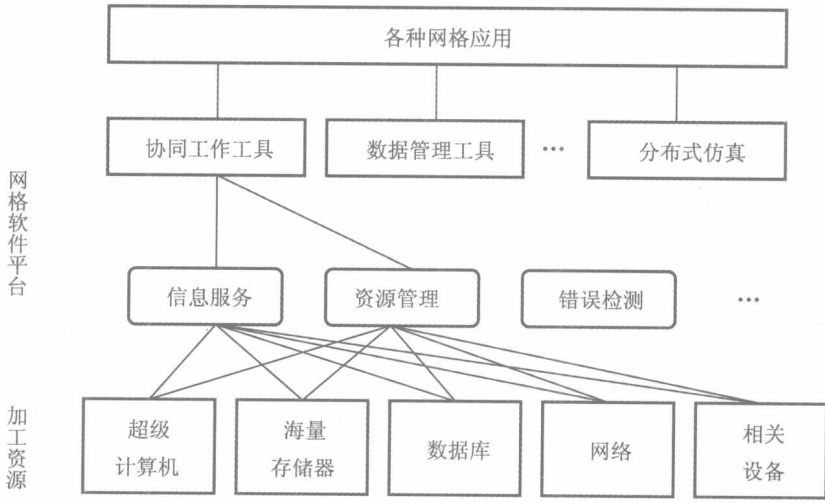


图 1-2 网格

网格和电力网的差别如表 1-1 所示。它们都有各自资源的提供者和消费者，对于电力网来说，资源提供者是发电站，消费者是各种消耗电能的设备；对于网格来说资源提供者是计算机等各类设备，消费者是使用网格计算能力进行问题求解的用户。网格和电力网都有覆盖范围广、组成资源多样和管理节点多的特点，但相对于电力网，网格的复杂性需要更高的技术要求。

表 1-1 网格与电力网

	网 格	电力网
节 点	计算中心	配送中心
生产设备	高性能计算机	发电机
加工资源	数据库、传感器、相关设备等	水能、火能、核能等
提供资源	计算能力、信息、知识	电能
传输网络	计算机网络	供电网络

(续表)

	网 格	电力网
控制系统	网格系统软件、中间件、网络缓存、负载均衡器等	电力调配系统
应用领域	科学计算、电子商务、信息服务等	动力、照明、多媒体等
应用设备	网格终端设备	各种电器
应用说明	统一界面	统一接口(插座、插头)

1.1.2 网格与网络

网格不同于网络,各类网络计算包括网格,虽然侧重点不同,但它们在资源共享、协同工作和开发应用等方面提供了良好的一致性。

1. 网络计算

网络计算中将各种由网络连接起来的自治资源和系统,以资源共享、协同工作和联合计算为手段,为各种用户提供基于网络的各类综合性服务。近年来出现的代表性模式包括企业计算、普及计算和对等计算等。

1) 企业计算(Enterprise Computing) 主要是指为实现企业级组织内部和组织之间的信息共享和协同工作而形成的网络计算模式,其核心是 C/S 模式和中间件技术。

人们在传统分布式系统基础上,开始采取中间件技术。20 世纪 90 年代,面向对象的中间件技术成为中间件平台的主流技术,技术热点是建立标准化的对象请求代理,屏蔽网络环境下的计算平台、操作系统、编程语言和网络协议等一大批各种层次上的异构性和复杂性,使得分布在网络环境下的各种应用系统能够协同工作,为各种网络应用提供管理服务和增值服务。其代表性技术有 SUN 公司的 EJB/J2EE、Microsoft 公司的 COM+/DNA 和 OMG 公司的 CORBE/OMA 等。

近年来,面向 Web 的企业计算解决方案成为解决企业信息共享和协同工作难点的首选。W3C(World Wide Web Consortium,万维网联盟)提出了 Web Services 技术体系,Microsoft 推出了 .Net 技术,SUN 公司推出了 SUN ONE 架构,企业计算技术进入一个新的时代。

2) 普及计算(Ubiquitous Computing, Pervasive Computing) 主要包括自然的人-机交互和网络计算,它强调人与计算环境的紧密联系,让计算机和网络更为有效地融入人们的生活,让人们更能方便地获得网络计算提供的各种服务。目前较为著名的有 MIT 的 Oxygen 研究计划、CMU 的 Aura 研究计划、UC Berkeley 的 Endeavour 计划和华盛顿大学的 Portolano 计划,这些计划的目标是提出全新的体系结构、应用模式、编程模型等基础理论模型和方法。

3) 对等计算(Peer to Peer,简称 P2P 模式) 是指网络上所有节点可以以平等的方式共享其他节点的计算资源。P2P 模式将网络计算模式从集中式计算引向分布式计算,网络应用从中央服务器向终端设备扩散,服务器与客户端的界限消失;服务器到服务器,服务器到客户端,客户端到客户端之间的所有设备都可建立起对话。正是这种模式给互联网带来了无尽的魅力,特别是在资源共享和文件交换以及协同工作和搜索引擎等方面,表现出强大优势。

2. 网格计算

网格计算是一种特殊的具有重大创新意义和发展潜力的分支计算网络。网格计算的最初目标是希望将计算机集成为一个能够进行远程控制的元计算机系统(Meta Computer),目前已经演化为将各种网络上的高性能计算机、信息系统、海量存储器、应用系统、模拟系统、虚拟

系统以及各种传感器和仪器设备集成在一起,将整个网络(主要是指 Internet 网)集成为一个功能强大的超级计算设施。

一个系统能否成为网格,不能仅仅看它的系统结构,而应该看系统能否为人们提供巨大的应用能力、商业价值和相关结论。2000 年,Ian Foster 在他的《网格剖析》一文中指出网格计算本质是在动态的、多机构的虚拟组织中协调资源共享和协同解决问题。2002 年,Ian Foster 又在《什么是网格?判断是否是网格的三个标准》一文中,又指出网格必须同时满足三个条件:

① 在非集中控制的环境中系统使用资源。网格能够整合各种资源,协调各种使用者,这些资源和使用者可以在不同控制域中;网格还必须解决在这种分布式环境中出现的安全、计费、权限等问题,否则就只能是一个本地管理系统。

② 使用标准、开放和通用的协议和接口。网格应用建立在多功能的协议和界面基础上,因此这些协议和界面接口必须经过认证、授权、资源开发和资源存取等环节,否则就只是一个具体应用系统。

③ 提供高品质的服务质量。网格通过协调多种资源,以得到多种服务,从而满足不同的消费者需要,如系统响应时间、流通量、有效性、安全性和资源重定位等,使系统的功效在原先各子系统功效基础上大大增强。

与电力网一样,网格设施作为国家的战略信息基础设施来规划、建设和运行管理,但它的复杂程度和难度远远大于电力网。这里面有技术难度问题,也有思想和观念问题,甚至是国家法律和政策层面的问题,需要经过多年努力才能不断完善。

与信息高速公路不同的是,网格属于处理信息的基础设施,而信息高速公路则是传输和获取信息的基础设施。网格是下一代真正意义上的信息基础设施,它将使以计算机为主体的信息处理发生根本的变化。

1.2 网格的产生与发展

网格是从连接超级计算点的项目开始的,当时这种方式被称为元计算,元计算这一术语是在美国 1989 年左右部署的千兆/秒的试验床项目 CASA 中出现的。网格计算的术语出现在 20 世纪 90 年代初的 I-WAY 项目。当时人们对计算机进行网络通信时形成的分布式系统,特别是它的设计、建造和部署等方面进行了大量的研究。人们通过中间件、库以及相关工具将分布的资源集成起来,形成统一的强有力的平台,以执行各种并行计算和分布式应用。其间出现了诸如元计算、可扩展计算(Scalable Computing)、全球计算(Global Computing)、因特网计算(Internet Computing)和目前的网格计算(Grid Computing)等术语。

Ian Foster 近年来又将网格问题定义为“在个人、组织和资源动态形成的集合体中灵活、安全地协同资源共享”。这个观点主要表达了信息的重要性,以及它对资源发现和互操作性是必须的。目前网格研究正从信息层面进入知识层面,同时这些与研究内容相关的 Web 技术和标准如同 XML 支持机器的通信、资源框架描述(Resource Description Framework, RDF)来表示元数据的交换等也得到了相应的发展。

网格的发展大致可以分为三个阶段:第一阶段指早期的网格计算;第二阶段指支持大规模数据与计算的中间件时期;第三个阶段指现在的重点在于分布式全球协作阶段,主要面向服务和信息层的问题。各阶段的划分对于找出各阶段的特征,有助于区分网格的问题、了解网格发

展起源,当然这一发展是连续的过程,它们之间的划分并没有明确的界限。

1.2.1 第一代网络

20世纪90年代初出现了早期的元计算和网格环境,其目的就是为广泛领域的高性能应用提供计算资源。其中,具有代表性的是 FAFNER 和 I-WAY 项目。

1. FAFNER 项目

1995年,由 Bellcore 实验室和 Syracuse 大学以及联合操作系统(Co-Operating System)共同主持的联盟启动了一个通过 Web 来进行因数分解的项目,它就是 FAFNER(Factoring via Network-Enabled Recursion)。FAFNER 项目用一种名为数字区分筛选(Number Field Sieve, NFS)的因数分解技术对 RSA130 进行因数分解,该技术需要用到计算 Web 服务器。

项目通过 Web 表格来激活服务器端的 CGI(Common Gateway Interface)脚本。系统可以支持多个筛选因数分解步骤的服务,包括 NSF 软件分布、项目文档、匿名用户注册、筛选任务的传播、关系集合、关系文档服务和实时筛选状态报告等。CGI 脚本支持机群管理,通过调整不同的睡眠时间,对单个筛选工作站进行管理,从而达到对机群拥有者影响最小的效果。贡献者可下载并建立一个 Web 客户端筛选软件守护程序,通过 HTTP 协议与 Web 服务器的 CGI 脚本进行交互,最后完成数据的计算任务。

FAFNER 系统最终在 1995 年的超级计算机会议(SC'95)中挑战 TeraFlop 活动中获奖。FAFNER 项目的成功原因首先在于采用了 NSF 软件分布技术,使得这种计算可以在普通的工作站(4MB 内存),以很小的边界值和很小的筛选片来开展有效的运算。其次,采用了匿名注册,有效保护了贡献者的身份。另外,通过本地 CGI 脚本运行,形成了 RSA130 Web 服务器的层次化网络,在吸收很多地方联盟的过程中,大大降低了潜在管理瓶颈,高效地开展筛选服务。

2. I-WAY 项目

I-WAY(Information Wide Area Year)项目是一个实验性的高性能网络,它连接了许多高性能计算机和高级可视化环境。它采用不同的路由与交换技术,将分布在全美 17 个站点的许多虚拟环境、数据集合和计算机连接起来。

在 I-WAY 项目中,使用了一个 I-POP(Installed Point of Presence)服务器用作 I-WAY 的网关。每一个加入 I-WAY 的站点都需要运行一个 I-POP 服务器。每个 I-POP 服务器实现统一的 I-WAY 认证、资源预留、进程创建以及通信等功能,同时它可以通过 Internet 访问,且在本站点的防火墙内进行有效操作。

I-WAY 项目的另一个主要角色是 I-Soft。I-Soft 是 I-POP 服务器在 UNIX 中采用的一种标准软件环境,它有助于解决异构、扩展性、计算性能以及安全等问题。

另外,在 I-WAY 项目中还开发了一个称之为计算资源代理 CRB(Computational Resource Broker)的资源调度器。CRB 的实现是通过单一集中调度器守护进程和多个局部调度器守护进程来构成的。每个 I-POP 服务器都有一个局部调度器守护进程。中心调度器维护了一个作业与表的队伍,用以表示局部机器的状态,为机器分配任务,在 AFS(Andrew File System)上维护状态信息等。

I-WAY 项目成功地应用在超级计算、访问远程资源、虚拟现实、视频以及 Web 和 G II 窗口等方面,并在 SC'95 中得到成功展示。

总之,无论是 FAFNER 项目还是 I-WAY 项目都试图建造一个元计算环境,只不过它们分别是从计算图谱的两个不同方向出发来集成资源的。FAFNER 系统是一个普适系统,可以在拥有 Web 服务器的任何平台上工作,它的客户机可以用一个低端的计算机充当;而 I-WAY 项目则是将超级计算中心的资源集成起来,但这两个项目都缺乏扩展性。

现在看来,FAFNER 项目存在一些明显不足,如需执行因数分解任务,每个客户端都需要重新编译连接,还需运行一个守护程序,且单个计算任务之间无法进行通信,也无法与它们的 Web 服务器进行通信。同样,在 I-WAY 项目中由于安装了 I-POP 平台,虽然用统一的方法来设置服务变得容易,但同时也意味着每个站点必须进行特殊安装才可以加入到 I-WAY 中。

FAFNER 项目和 I-WAY 项目的实施,对第二代网格项目的成功具有十分重要的参考价值。可以这样说,FAFNER 项目对类似于 SET@home 和分布式 Net 技术的产生,I-WAY 项目对 Globus 和 Legion 的产生都具有十分巨大的影响。

1.2.2 第二代网格

第一代网格项目在超级计算中心的连接上取得了成功,人们在此基础上取得了一些重要的技术突破,采用更高、带宽更大的网络技术,建立网格标准,以支持大规模计算和大量数据处理的多种应用,将网格看作是全局范围的更具活力的分布式基础设施,使网格更具普适性,我们把这种网格称为第二代网格。

在第二代网格建设中重点是需解决异构性、扩展性和适应性等三个问题。

① 异构性是指网格建设中会涉及多种多样的资源,这些资源在本质上是不同的,会跨越很多管理范围,甚至遍布全球范围。

② 扩展性是指网格资源可能在很短时间内会从较少的资源迅速扩展到成千上万的各种资源。如何支持大量分布资源的应用,克服通信延迟,提高访问资源的局部性;如何更进一步解决由于资源规模的进一步扩大而带来更多的管理组织问题,以及异构问题、认证问题和信任问题等;更大规模的应用可有其他应用组成,这从一定程度上又更增加了系统的“智力复杂度 (Intellectual Complexity)”。

③ 适应性是指网格上资源的多样性也容易导致差错的可能性,资源管理者和资源应用应根据资源情况动态调整它们的行为,以便从可用资源与服务中获取最大的性能。

解决异构问题的关键是建立并使用标准。由于许多系统使用不同的标准和系统应用程序接口,这就需要在网格环境中,将大量的服务与应用移植到大量不同的计算机系统上。建立相同的信息交换格式有助于降低复杂性。在网格中,中间件通过提供一系列标准的服务接口,可以隐蔽资源的异构特性,为用户以及应用提供一个同构和无缝的环境。近年来,中间件的重新兴起,作为一种集成的软件应用运行在分布的异构环境上。中间件常常可看作是操作系统和应用之间的软件层,为确保应用的执行提供各种服务。

在第二代网格中具有代表性的是 Globus 项目和 Legion 项目。

Globus 作为一个软件基础设施平台,应用可以将分布的异构计算资源作为一个虚拟的计算机来使用。Globus 同时可以提供可靠、一致、随处可以访问的高端计算能力,可以屏蔽掉由于资源与用户的分布所带来的影响。Globus 系统的核心组成部分是 Globus 工具包,工具包提供了建造计算网格所需要的基本服务与能力。该工具包有多个功能强大的组件组成,实现

了安全、资源定位、资源管理和通信等基本服务。Globus 的重点已经从原先的仅仅支持高性能应用,转到可以对虚拟组织进行支持的普适服务。

Legion 是一个基于对象的“元系统(Metasytem)”,通过它可以实现异构、地理分布的高性能计算系统的无缝交互。Legion 试图通过用户工作站,屏蔽掉规模、物理位置、语言和底层特定操作系统的不同,为用户提供一个集成的基础设施。Legion 提供了不同于 Globus 的网格环境功能,它将所有的组件都封装成对象,具备了所有常用面向对象方法的优点。

目前在第二代网格系统中,网格的核心软件,已经从一些较为超前的为大型计算密集型高性能应用提供专门服务的项目(如 Globus GT1、Legion)基础上呈现出更为通用、更为开放的趋势,比如 Globus GT2 和 Avaki。同时随着核心软件发展的大量的工具和实用程序的开发,为用户和服务提供了更高更好的服务,对等式技术就是在此时出现的。

1.2.3 第三代网格

第二代网格提供的互操作性是大规模计算所必需的。随着网格方案的深入开发与发展,网格工程的其他方面应用、需求以及技术的发展也变得更加清晰起来。为建造更新的网格应用,重组已有的组件和信息资源,将这些组件以灵活的方式组合。这与第三代网格的面向服务和对元数据关注有密切关系。

在第三代网格系统中,其主要特点是面向服务的体系结构。早在 2001 年,多种网格项目就出现了许多网格体系项目。在 Ian Foster 的《网格解剖》一文中就提出了层次化的模型,而在 IPG(Information Power Grid)项目中将一些服务形成各种集合,并没有按层进行组织。此时,Web 服务模型也流行起来,逐步成为一种支持面向服务方法的很有前途的标准。

同时,第三代网格系统一改第二代网格所重点描述的大规模数据处理与计算,特别强调了“分布式协作”和“虚拟组织”两个概念。第三代网格对网格计算具有更全局的观点,更加强调需求而不是强调具体的支撑技术。同时,在第三代网格系统中,更加强调自动化思想,当人们无法处理大规模和各种异构性带来的难题时,系统引入自治机制将难题交给进程来自动处理。

第三代网格的典型代表是万维网服务和开放网格服务结构。万维网服务是面向商业领域的计算框架。作为万维网服务框架中的基本技术,简单对象访问协议(SOAP)、万维网服务描述语言(WSDL)和统一集成、描述和发现(UDDI)规范获得业界的普遍支持。2002 年发布的开放网格服务结构(OGSA)则把万维网服务和网格计算结合起来,建立了以服务为核心的网格结构。

1.2.4 网格的研究现状

网格的研究经过前一时期的发展,已经从美国推广到世界各大洲,各个国家和地区都投入了大量的资金进行网格技术研究和网格基础设施建设。网格技术的出现使许多行业富有挑战性的问题在网格这一新的计算平台上得以解决。各国研究机构和商业机构纷纷在网格研究上投入巨资开展研究和基础设施建设,各个国际和地区的研究机构和商业企业之间的有效合作也逐步开展起来。各种网格论坛、网格协作组织纷纷出现,其中最著名的全球最大的网格联盟——全球网格论坛(Global Grid Forum,GGF),是一个网格技术论坛的国际组织,由网格论坛和欧洲网格论坛以及亚太网格社区组成,已经成为事实上的网格标准的制定、发布和推动的应用机构。

1. 美国网格研究

网格技术研究起源于美国,美国也一直是世界网格技术研究和网格基础设施建设的领头羊。多年来,美国在网格研究上投入了大量的人力和物力,多家研究机构开展了与网格相关的工作,先后开发了 Condor、Legion、Globus、DOCT、PUNCH 和 Entropia PC 等多个有影响的软件与工具。

许多大公司也纷纷加入到网格技术的研究中,宣布本公司的网格战略计划。IBM 和 Globus 提出的开放网格服务机构正在赢得大家的认可,并将对未来网格的研究产生重大的影响。

同时,美国国家科学基金也于 2001 年发起一个创新计划,投资 1 200 万美元,计划用 3 时间来构建和部署先进的网格服务,简化对国际互联网上信息和服务的访问。专门向科研人员和教育界人士提供包括 Globus、CondorG 和 NWS 等中间件包,来推动网格技术的研究和应用。

2. 欧洲网格研究

欧洲作为目前世界网格技术研究最活跃的地区之一,已经先后启动了多个网格研究项目,其中影响较大的有欧洲网格和数据网格项目,其他各国也有自己的网格研究项目。

欧洲网格主要是为科学研究服务的,旨在建立一个给用户的安全、简单、透明访问欧洲范围内信息资源的平台。目前已经由包括德国、英国、法国、瑞士、挪威和波兰等国家的大学和计算中心参加,另外还有一些航空、天文和高能物理等应用领域的机构和组织参加,主要开展网格技术研究、网格基础设施建设和应用开发等三方面工作。

欧洲数据网格(European Data Grid, EDG)是欧盟支持的一个项目,主要目标是建立一个下一代计算基础设施,通过该项目提供高强度、共享数据达 10^{14} B 的分布式数据库资源。该项目由包括中间件、基础设施、应用和管理等四个方面的 12 个工作组组成。目前参加成员除欧盟外还有美国、日本、俄罗斯、希腊、波兰和塞浦路斯等多个国家和地区的研究人员。

UNICORE 是德国联邦教育和研究部资助的一个研究项目。其目的是为用户提供一套软件,允许用户向远程高性能计算机提交自己的作业,而不需要知道远程计算机的操作系统、数据存储格式、管理策略和过程。它最大可能地结合了当前的 Web 技术、JAVA 用户接口语言和相应的浏览器工具。用户可以通过授权访问 UNICORE 资源。目前,该项目有多家德国研究机构的合作者和多个包括中等范围天气预报欧洲中心、莱布尼兹计算中心、帕德博恩并行计算中心、HP、IBM 等多家公司。

另外,还有荷兰网格、爱尔兰网格,以及由挪威、丹麦、瑞典和芬兰四国参加的 Nordu 网格。

3. 我国网格研究

网格研究在中国已经列入“863 计划”。1996 年中科院计算技术研究所开始网格技术的研究开发工作,承担了“863 计划”重点项目“国家高性能计算环境(NHPCE)”,2000 年开发了连接国内 8 个曙光计算中心节点的网格。2001 年提出了“织女星网格”计划,计划从网格操作系统、信息网格和知识网格三个层次进行研究。2002 年,科技部召开了“网格战略研讨会”,将网格的研究和应用列为“863 计划”的一个专项,并准备投入 3 亿元资金建立“中国国家网格”(China National Grid, CNGrid),主要研制面向网格的亿万次级高性能计算机、具有数万亿次聚合计算能力的高性能计算环境;开发具有自主知识产权的网格软件;建设科学研究、经济建设、社会发展和国防建设急需的重要应用网格;制定若干与网格相关的国家标准,参与国际标

准的制定。目前,“中国国家网格”CNGrid已经在全国设立多个网格节点,其中放置在上海超级计算中心的曙光 4000A 的速度达 10 亿次,放置在北京中科院计算技术研究所网络信息中心的联想深腾 6800 的计算速度达到 4 万亿次。2005 年,中国国家网格运行管理中心在中科院计算技术研究所正式揭牌,标志着中国国家网格的正式开通运行。

教育部于 1999 年支持清华大学网格研究组进行了先进计算基础设施(Advanced Computational Infrastructure, ACI)项目研究;教育部于 2003 年与 IBM 合作投资 2 000 万元共建“中国教育科学网格项目”(China Grid),全国共有 12 所高校参加。该项目实施目的是通过网格技术将全国 211 建设的 100 所重点大学的资源广泛共享,并争取我国在网格计算的基础研究和应用研究方面走在世界前列。IBM 为该项目提供网格中间件、服务器以及全面存储解决方案。该项目计划根据开放式网格服务结构和网格服务等标准建设一个中国网格应用联合研发中心。

1.3 网格的特点

网格是建立在因特网和 Web 基础上的新一代技术,传统因特网实现了计算机网络的连通,将各个独立的计算机个体联结成网络,但并没办法共享其他机器的资源;Web 的兴起,实现了网页的连通,它可以实施包括电子商务领域的众多功能,但各个行业在应用层面上的互联互通仍然无法实现。网格技术友好地实现了互联网上的包括计算资源、存储资源、通信资源、软件资源、信息资源和知识资源等所有资源应用层面上的全面共享和协同工作。

网格并不能替代因特网。网格是构筑在因特网上的一组新兴技术,它比当前的因特网技术性能更高、功能更强、应用更广,能让人们更加透明地使用计算、存储等其他资源。网格所采用的新的共享方案并不意味着取代现有的因特网技术,与网络相比,网络是网格的一个很重要的基础,网格是在现有网络技术上建立的更高层次、更全面的资源共享,其计算技术也更加先进。网格的优秀特点可以描述如下:

1) 高性能 网格的高性能是指它在计算速度、数据处理速度、带宽和资源利用上有更加明显的优势。网格通过网格的中间件软件将分布在不同地点的高性能的超级计算机聚合起来,形成超高速的计算平台,为企业提供当前因特网所无法提供的大量的高性能计算和数据处理速度;其次,网格计算需要更高要求的传输带宽;同时,网格采用一种“广域缓存技术”,自动把用户最需要的信息放在离用户最近的服务器上,更加有效地利用网络中的资源。

2) 分布性 网格的资源包括计算资源、数据资源、存储资源和仪器设备资源,它们分布在不同地理位置,而网格工作流关键往往不是集中在一起的。因此在这种分布式环境下,需要解决网格资源的多个问题,主要包括任务分配和调度问题、传输与通信问题、人与系统及人与人之间的交互和协同问题以及网格应用的自动执行和协作等问题。

3) 异构性 网格的资源具有异构性,其组成是多种多样的,正是这种异构性才形成网格资源的丰富性。就计算资源而言,有不同类型的计算机,不同的计算方式,不同的计算接口,不同的系统架构;同样,对于存储资源和数据资源而言,也必然会有各种不同的资源。正是网格这种处理资源异构性的能力,实现了网格具有良好的资源管理能力。

4) 自治性 网格资源的多样性还表现在不同的资源拥有者,而资源属于不同的资源拥有者也决定了资源管理上的自治性,网格资源拥有者拥有资源的最高管理权限。同时,这些资源