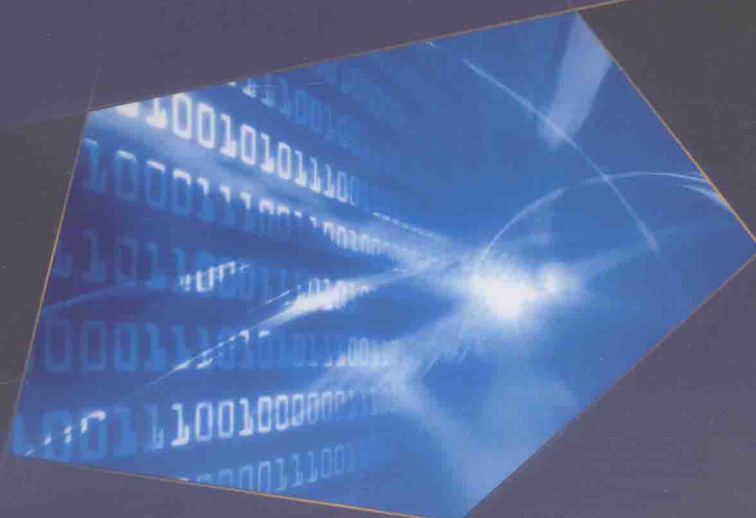


以网络为基础的
科学活动环境研究系列

网络计算环境： 体系结构

单志广 姜进磊 武永卫 著



科学出版社

以网络为基础的科学活动环境研究系列

网络计算环境：体系结构

单志广 姜进磊 武永卫 著

科学出版社

北 京

内 容 简 介

网格技术作为支撑现代科学活动的重要手段,受到了人们的普遍重视。然而,构造一个实用的网格系统并不是一件容易的事情,面临诸多的技术挑战。本书试图从体系结构的角度来描述一个网格系统所应该具有的各种元素及所提供的功能,为开发者提供参考。本书首先介绍了网格的三种典型架构,即五层沙漏结构、开放网格服务架构和面向服务的网格体系。其次,从理论的角度探讨了不同网格系统的统一描述。最后从实现的角度,对元信息服务、监控服务、数据管理、执行管理及安全服务等网格系统的关键部件进行了剖析。

本书可供网络计算相关领域的科研技术人员参考,也可供相关专业的研究生、本科生阅读。

图书在版编目(CIP)数据

网络计算环境:体系结构/单志广,姜进磊,武永卫著. —北京:科学出版社,2014.10

(以网络为基础的科学活动环境研究系列)

ISBN 978-7-03-042158-6

I. ①网… II. ①单… ②姜… ③武… III. ①网络环境 IV. ①TP393

中国版本图书馆 CIP 数据核字(2014)第 237989 号

责任编辑:任 静 / 责任校对:朱光兰

责任印制:徐晓晨 / 封面设计:迷底书装

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京京华虎彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

*

2014 年 10 月第 一 版 开本:720×1 000 1/16

2014 年 10 月第一次印刷 印张:10 1/2

字数:192 000

定价:65.00 元

(如有印装质量问题,我社负责调换)

“以网络为基础的科学活动环境研究系列”编委会

编委会主任：

何新贵，研究员，中国工程院院士，北京大学信息科学技术学院

编委会委员（按姓氏笔画为序）：

王 斌，研究员，中国科学院大气物理研究所

白晓民，研究员，中国电力科学研究院

杨义先，教授，北京邮电大学计算机学院

陈和生，研究员，中国科学院院士，中国科学院高能物理研究所

陈润生，研究员，中国科学院院士，中国科学院生物物理研究所

郑纬民，教授，清华大学计算机科学与技术系

单志广，研究员，国家信息中心

桂亚东，研究员，上海超级计算中心

钱德沛，教授，北京航空航天大学计算机学院

徐志伟，研究员，中国科学院计算技术研究所

序

近年来,以网络为基础的科学活动环境已经引起了各国政府、学术界和工业界的高度重视,各国政府纷纷立项对网络计算环境进行研究和开发。我国在这一领域同样具有重大的应用需求,同时也具备了一定的研究基础。以网络为基础的科学活动环境研究将为高能物理、大气、天文、生物信息等许多重大应用领域提供科学活动的虚拟计算环境,必然将对我国社会和经济的发展、国防、科学研究,以及人们的生活和工作方式产生巨大的影响。

以网络为基础的科学活动环境是利用网络技术将地理上位置不同的计算设施、存储设备、仪器仪表等集成在一起,建立大规模计算和数据处理的通用基础支撑结构,实现互联网上计算资源、数据资源和服务资源的广泛共享、有效聚合和充分释放,从而建立一个能够实现区域或全球合作或协作的虚拟科研和实验环境,支持以大规模计算和数据处理为特征的科学活动,改变和提高目前科学研究工作的方式与效率。

目前,网络计算的发展基本上还处于初始阶段,发展动力主要来源于“需求牵引”,在基础理论和关键技术等方面的研究仍面临着一系列根本性挑战。以网络为基础的科学活动环境的主要特性包括:

(1)无序成长性。**Internet** 上的资源急剧膨胀,其相互关联关系不断发生变化,缺乏有效的组织与管理,呈现出无序成长的状态,使得人们已经很难有效地控制整个网络系统。

(2)局部自治性。**Internet** 上的局部自治系统各自为政,相互之间缺乏有效的交互、协作和协同能力,难以联合起来共同完成大型的应用任务,严重影响了全系统综合效用的发挥,也影响了局部系统的利用率。

(3)资源异构性。**Internet** 上的各种软件/硬件资源存在着多方面的差异,这种千差万别的状态影响了网络计算系统的可扩展性,加大了网络计算系统的使用难度,在一定程度上限制了网络计算的发展空间。

(4)海量信息共享复杂性。在很多科学研究活动中往往会得到 PB 数量级的海量数据。由于 **Internet** 上信息的存储缺少结构性,信息又有形态、时态的形式多样化的特点,这种分布的、半结构化的、多样化的信息造成了海量信息系统中信息广泛共享的复杂性。

鉴于人们对于网络计算的模型、方法和技术等问题的认识还比较肤浅,基于 **Internet** 的网络计算环境的基础研究还十分缺乏,以网络为基础的科学活动环境还存在着许多重大的基础科学问题需要解决,主要包括:

(1) 无序成长性与动态有序性的统一。Internet 是一个无集中控制的不断无序成长的系统。这种成长性表现为 Internet 覆盖的地域不断扩大, 大量分布的异构的资源不断更新与扩展, 各局部自治系统之间的关联关系不断动态变化, 使用 Internet 的人群越来越广泛, 进入 Internet 的方式不断丰富。如何在一个不断无序成长的网络计算环境中, 为用户任务确定所需的资源集合, 进行动态有序的组织和管理, 保证所需资源及其关联关系的相对稳定, 建立相对稳定的计算系统视图, 这是实现网络计算环境的重要前提。

(2) 自治条件下的协同性与安全保证。Internet 是由众多局部自治系统构成的大系统。这些局部自治系统能够在自身的局部视图下控制自己的行为, 为各自的用户提供服务, 但它们缺乏与其他系统协同工作的能力及安全保障机制, 尤其是与跨领域系统的协同工作能力与安全保障。针对系统的局部自治性, 如何建立多个系统资源之间的关联关系, 保持系统资源之间共享关系定义的灵活性和资源共享的高度可控性, 如何在多个层次上实现局部自治系统之间的协同工作与群组安全, 这些都是实现网络计算环境的核心问题。

(3) 异构环境下的系统可用性和易用性。Internet 中的各种资源存在着形态、性能、功能, 以及使用和服务方式等多个方面的差异, 这种多层次的异构性和系统状态的不确定性造成了用户有效使用系统各种资源的巨大困难。在网络计算环境中, 如何准确简便地使用程序设计语言等方式描述应用问题和资源需求, 如何使软件系统能够适应异构动态变化的环境, 保证网络计算系统的可用性、易用性和可靠性, 使用户能够便捷有效地开发和使用系统聚合的效能, 是实现网络计算环境的关键问题。

(4) 海量信息的结构化组织与管理。Internet 上的信息与数据资源是海量的, 各个资源之间基本上都是孤立的, 没有实现有效的融合。在网络计算环境下如何实现高效的数据传输, 如何有效地分配和存储数据以满足上层应用对于数据存取的需求, 以及有效的数据管理模式与机制, 这些都是网络计算环境中数据处理所面临的核心问题。为此需要研究数据存储的结构和方法, 研究由多个存储系统组成的网络存储系统的统一视图和统一访问, 数据的缓冲存储技术等海量信息的组织与管理方法。

为此, 国家自然科学基金委员会于 2003 年启动了“以网络为基础的科学活动环境研究”重大研究计划, 着力开展网络计算环境的基础科学理论、体系结构与核心技术、综合试验平台三个层次中的基本科学问题和关键技术研究, 同时重点建立高能物理、大气信息等网络计算环境实验应用系统, 以网络计算环境中所涉及的新理论、新结构、新方法和新技术为突破口, 力图在科学理论和实验技术方面实现源头创新, 提高我国在网络计算环境领域的整体创新能力和国际竞争力。

在“以网络为基础的科学活动环境研究”重大研究计划执行过程中, 学术指导专家组注重以网络标准规范研究作为重要抓手, 整合重大研究计划的优势研究队伍,

推动集成、深化和提升该重大研究计划已有成果，促进学术团队的互动融合、技术方法的标准固化、研究成果的集成升华。在学术指导专家组的研究和提议下，该重大研究计划于 2009 年专门设立和启动了“网格标准基础研究”专项集成性项目(No.90812001)，基于重大研究计划的前期研究积累，整合了国内相关国家级网格项目平台的核心研制单位和优势研究团队，在学术指导专家组的指导下，重点开展了网格术语、网格标准的制定机制、网格标准的统一表示和形式化描述方法、网格系统结构、网格功能模块分解、模块内部运行机制和内外接口定义等方面的基础研究，形成了《网格标准的基础研究与框架》专题研究报告，研究并编制完成了网格体系结构标准、网格资源描述标准、网格服务元信息管理规范、网格数据管理接口规范、网格互操作框架、网格计算系统管理框架、网格 workflow 规范、网格监控系统参考模型、网络安全技术标准、结构化数据整合、应用部署接口框架(ADIF)、网格服务调试结构及接口等十二项网格标准研究草案，其中两项已列入国家标准计划，四项提为国家标准建议，十项经重大研究计划指导专家组评审成为专家组推荐标准，形成了描述类、操作类、应用类、安全保密类和管理类五大类统一规范的网格标准体系草案，相关标准研究成果已在我国三大网格平台 CGSP、GOS、CROWN 中得到初步应用，成为我国首个整体性网格标准草案的基础研究和制定工作。

本套丛书源自“网格标准基础研究”专项集成性项目的相关研究成果，主要从网络计算环境的体系结构、数据管理、资源管理与互操作、应用开发与部署四个方面，系统展示了相关研究成果和工作进展。相信本套丛书的出版，将对于提升网络计算环境的基础研究水平、规范网格系统的实现和应用、增强我国在网络计算环境基础研究和标准规范制订方面的国际影响力具有重要的意义。

是以以为序。



北京大学教授

国家自然科学基金委员会“以网络为基础的科学活动环境研究”

重大研究计划学术指导专家组组长

2014 年 10 月

前 言

计算技术已经深刻地改变了我们的工作、学习和生活，成为继理论和实验之后人类认识世界和改造世界的第三大利器，信息和通信技术的飞速发展和进一步融合，以及体系结构的进步更是将其提升到前所未有的高度。

科学活动是人类的一种认识和实践活动，其最基本的特点是创造性和探索性。自 20 世纪初的物理学革命开始，科学活动步入了现代化的阶段：一方面，科学研究的领域和对象逐渐向微观和宏观各层次深入；另一方面，学科之间的横向和纵向联系变得更加紧密。除此之外，现代科学活动还日益表现出技术化、产业化、社会化和全球化的趋势。这些特点越来越需要知识和技术的结合、相关学科信息的共享和多种科研资源的协同工作。

网格，也称为网格计算，是为解决跨组织、跨地域的大规模资源共享和协作而提出的一种新方法。它被定义为在一个分布式、异构的环境下对跨传统的管理域和组织域的资源进行虚拟化和集成，对服务进行管理的一类分布式系统。从英国的 e-Science 项目开始，网格成为支撑现代科学活动的核心技术，获得了蓬勃的发展。

虽然国内外的网格研究已经取得了重大的进展，并开始投入实际应用，但各网格计划的研究都相对独立，造成各网格平台之间的互联互通和互操作性较差，网格之间的资源较难进行统一的监控和调度，从而无法实现跨平台的资源共享。而且，在不同的网格研究项目中，解决的基本核心问题都非常类似，从世界范围看，网格研究重复投资、重复开发的现象严重，造成了人力和物力的大量浪费，在很大程度上阻碍了网格技术的发展。因此，实现异构网格之间的互通互操作具有非常重要的理论和现实意义。

标准体系是网格应用发展的重要前提，也是实现异构网格之间的互通互操作的基础。如果没有标准体系作为支撑，就无法真正实现网格系统的互联互通、资源共享和协同工作。目前国际上虽然已经推出了一些网格标准，但大多处于初级阶段，还没有得到业界的广泛认可。尽管如此，在核心技术上，相关机构与企业已达成共识：由美国阿贡 (Argonne) 国家实验室与南加州大学信息科学学院 (ISI) 合作开发的 Globus Toolkit (GT) 已成为网格计算事实上的标准。GT 提供了构建网格应用所需的很多基本服务，如安全、资源发现、资源管理、数据访问等。

另外，2002 年 2 月，在加拿大多伦多市召开的全球网格论坛 (Global Grid Forum, GGF) 会议上，Globus 项目组和 IBM 共同倡议了一个全新的网格标准 OGSA。OGSA 叫作开放网格服务架构 (Open Grid Services Architecture)，它把 Globus 标准与以商

用为主的 Web Service 的标准结合起来, 网格服务统一以 Service 的方式对外界提供。以 OGSA 的提出为标志, 结合了 Web 服务技术的网格技术的标准化工作不仅受到了 OMG、W3C、OASIS 等已有国际标准化组织的关注和支持, 网格领域中也先后出现了 GGF、EGA、OGF 等标准化组织。OGSA 的诞生, 标志着网格已经从学术界延伸到了商业界, 而且从一个封闭的环境走向开放的世界中。

尽管 OGSA 从一诞生就得到业界的广泛支持, 为众多的国际知名企业和研究机构所接受, 但 OGSA 只是概念模型, 定义了网格概念结构, 没有涉及任何实现层面和功能接口层面描述, 这就导致目前网格各种异构的实现虽然都遵循了 OGSA 标准, 但无法实现真正意义上的交互, 也就不能实现真正意义上的资源整合和计算协同。此外, OGSA 是基于国际上对网格的统一认识和研究结果而提出的, 缺乏对中国具体国情和应用环境的特性考虑, 因此制定我国自己的网格体系结构标准对国家的网格系统的研发、提高我国网格行业的产出质量、提升我国网格技术的竞争力, 以及力争与国外不但在网格技术研究而且在网格标准规范制定方面达到同等对话层次等方面都具有重要的意义。

在上述背景下, 我们对网格的相关标准进行了研究, 内容涉及网格体系结构、网格资源描述、网格服务元信息管理、网格数据管理、网格计算系统管理、网格工作流、网格监控、网络安全、网格互操作等多个层面。本书是网格体系结构研究成果的集中体现, 参与研究的单位包括国家信息中心、清华大学、中国电子技术标准化研究所、北京航空航天大学、中国科学院计算技术研究所、中国科学院软件研究所、北京大学、兰州大学、北京邮电大学等。

本书作者们的研究工作得到了国家自然科学基金项目“网格标准基础研究”(No.90812001)的资助, 并得到了国家自然科学基金委员会“以网络为基础的科学活动环境研究”重大研究计划学术指导专家组的悉心指导, 在此表示深深的谢意!

在编著本书的过程中, 我们力求内容翔实准确, 但限于水平和时间, 疏漏之处在所难免, 敬请读者批评指正。

作 者

2014 年 8 月

目 录

前言

第 1 章 现代科学活动环境与网格技术综述	1
1.1 现代科学活动的特点及要求	1
1.2 网格概念的提出与发展	2
1.3 网格的特点	3
1.4 网格在国内外的发展现状	4
1.5 网格面临的挑战	6
1.6 与网格有关的标准和组织	8
参考文献	9
第 2 章 五层沙漏结构	11
2.1 网格体系结构	11
2.2 沙漏结构的基本理念	12
2.3 沙漏结构详述	15
2.3.1 构造层	15
2.3.2 连接层	16
2.3.3 资源层	16
2.3.4 汇聚层	17
2.3.5 应用层	17
参考文献	18
第 3 章 开放网格服务架构	19
3.1 SOA	19
3.1.1 概念辨识	19
3.1.2 架构描述	20
3.1.3 服务的类型与接口	20
3.1.4 SOA 的特点	22
3.2 Web 服务	23
3.2.1 SOAP	23
3.2.2 WSDL	25

3.2.3	UDDI	26
3.3	OGSA 的基本理念	27
3.4	结构示意	28
3.4.1	资源层	28
3.4.2	Web 服务层	29
3.4.3	OGSA 架构服务层	30
3.4.4	应用程序层	31
3.5	OGSA 中的服务接口	31
3.6	OGSA 的核心服务	32
3.6.1	执行管理服务	32
3.6.2	数据服务	33
3.6.3	资源管理服务	33
3.6.4	安全服务	34
3.6.5	自我管理服务	34
3.6.6	信息服务	35
	参考文献	35
第 4 章	面向服务的网格体系	37
4.1	基本概念	37
4.1.1	以服务为中心	37
4.1.2	采用统一的 Web 服务框架	38
4.2	层次结构	39
4.2.1	物理和逻辑资源层	39
4.2.2	Web 服务适配层	39
4.2.3	标准网格服务层	39
4.2.4	网格应用层	40
4.3	运行环境	40
4.3.1	简单运行环境	40
4.3.2	虚拟运行环境	40
4.3.3	组操作环境	41
4.4	网格的功能模块	41
4.5	网格的执行流程	43
4.5.1	作业的类型	43
4.5.2	作业的执行模型	43
4.5.3	作业的执行流程	45

4.6 参考实现: CGSP	47
参考文献	48
第 5 章 数学基础	50
5.1 实体与本体	50
5.2 虚拟化与代数系统	50
5.3 类型代数与商代数	52
5.4 同型与异构性	54
第 6 章 网格标准的理论描述	56
6.1 描述要点	56
6.1.1 资源相关的标准	57
6.1.2 工作流相关的标准	58
6.1.3 日志相关的规范	64
6.1.4 通知与资源的动态管理	65
6.2 握手与通道	65
6.3 本地代理	66
6.4 移动性	67
6.5 程序设计	67
6.6 与外网的连接	68
第 7 章 元信息服务	70
7.1 概述	70
7.2 功能架构	71
7.2.1 总体描述	71
7.2.2 MDS	72
7.2.3 MCAT	73
7.2.4 MCS	74
7.3 模块描述	76
7.3.1 元信息提供者及信息类型	76
7.3.2 元信息服务模型	76
7.3.3 元信息服务组件	79
参考文献	81
第 8 章 监控服务	82
8.1 概述	82
8.2 技术难点	83

8.3 常用监控模型与系统分类	84
8.3.1 常用监控模型	84
8.3.2 监控系统分类	87
8.4 监控系统的标准架构	88
8.5 网格监控标准的参考实现：CGSV	90
参考文献	92
第 9 章 数据管理	94
9.1 概述	94
9.2 GT4 中的数据管理	95
9.3 关键技术	96
9.3.1 GridFTP	96
9.3.2 RFT	99
9.3.3 RLS	101
9.3.4 DAI	102
9.4 其他数据管理系统	107
9.4.1 SRM	107
9.4.2 SRB	109
9.5 数据管理技术应用示例	110
参考文献	113
第 10 章 执行管理	115
10.1 概述	115
10.2 执行管理剖析	116
10.3 作业需求描述	118
10.4 资源分配与作业调度	121
10.4.1 目标	121
10.4.2 调度模型	122
10.4.3 资源分配模型	123
10.5 典型的执行管理系统	127
10.5.1 GRAM	127
10.5.2 GARA	128
10.5.3 Condor	130
10.5.4 GRACE	131
10.5.5 CGSP 的执行管理	132
参考文献	135

第 11 章 安全服务	137
11.1 概述	137
11.2 安全技术基础	138
11.2.1 PKI	138
11.2.2 TLS	140
11.2.3 SAML	141
11.2.4 Web 服务安全	143
11.3 网络安全基础设施	143
11.3.1 GSI 概览	143
11.3.2 相互认证	144
11.3.3 安全通信	145
11.3.4 授权	145
11.3.5 委托与单点登录	146
11.4 GOS 安全机制	147
11.4.1 基本概念	147
11.4.2 安全体系	147
11.4.3 认证的实现	148
11.4.4 授权的实现	149
11.4.5 通信安全的实现	150
参考文献	150

现代科学活动环境与网格技术综述

从英国的 e-Science 项目开始, 网格成为支撑现代科学活动的核心技术, 获得了蓬勃的发展。本章从现代科学活动的特点及要求出发, 对网格技术的历史、发展现状、面临的挑战等作一个概括而又全方位的介绍。

1.1 现代科学活动的特点及要求

科学研究是人类的一种认识和实践活动, 其最基本的特点是创造性和探索性。现代科学活动始于 20 世纪初的物理学革命, 其突出特点表现在以下八个方面^[1]。

(1) 科学研究的领域和对象逐渐向微观和宏观各层次深入, 对过程、结构和功能多个方面进行完整研究。

(2) 科学研究的内容具有学科交叉的性质, 学科之间的横向和纵向联系更加紧密。

(3) 科学研究的组织形式更加多样, 集体研究成为科学研究的主要形式, 科学研究成为一种重要的社会职业部门或社会建制, 科研人员的数目剧增, 形成庞大的科研队伍。

(4) 科学研究的方法和手段越来越依赖于最新的复杂技术装备, 信息技术、网络技术等广泛应用于科学研究的各个领域, 呈现出科学技术化的趋势。

(5) 科学研究的成果迅速转化和扩散, 使得基础研究、应用研究和开发研究三者之间的界限越来越模糊, 政府、企业和科研机构之间的关系更加紧密, 呈现出科学产业化的趋势。

(6) 科学技术已经渗透到社会生活的各个领域, 更加强调与国家经济、安全和可持续发展的目标紧密结合; 同时科学研究的实验设备日益庞大和昂贵, 对社会的人力、物力和资金的需求也不断加大, 呈现出科学社会化的趋势。

(7) 科技资源的配置在全球范围内进行, 科技成果的评价和应用也在全球范围内进行和流动, 国际科技合作与交流迅速增加, 科学研究呈现出全球化的特征。

(8) 科学研究的质量控制既要关注研究工作的潜在应用, 更要考虑研究成果的

可使用性、成本效益和社会可接受性等，而且更加强调科学家的社会责任，并把伦理标准纳入科学研究的行为规范之中。

现代科学活动的这些特点越来越需要知识和技术的结合、相关学科信息的共享和多种科研资源的协同工作。因此，如何协同分散在各地的大量科研资源来完成各种复杂科研问题的求解已成为一个至关重要的问题，于是网格技术应运而生。

1.2 网格概念的提出与发展

网格(Grid)技术正是为解决跨组织、跨地域的大规模资源共享和协作问题而提出的一种新方法。网格最早是借助电力网的概念提出的：就像人们使用电力而不用知道电力从哪里来、怎么来一样，人们在使用网格提供的计算力的时候也无需知道提供“计算力”的资源的位置、互联方式等细节问题。网格问题被形象地定义为在个人、组织机构、互联资源(计算设备、网络、在线仪器设备、存储设备等)的动态集合上实现灵活、安全、透明、协同的资源共享。网格研究试图将一组通过高速网络连接起来的异构的资源聚合起来，作为一个整体计算环境，透明地向用户提供各类高性能计算服务。其最终目标是希望计算机一旦接入网络就能获取源源不断的计算能力。

从狭义上来讲，网格被称为计算网格(Computational Grid)，由元计算(Metacomputing)^[2]的概念发展而来。在刚开始的时候，元计算的目标是把各个独立的超级计算机或网络上的闲置计算机资源集合起来，成为一个整体来为科学计算提供强大的超级计算服务。随着网络环境及网络应用需求的多样化，单纯的超级计算服务已经无法满足这种需求，在这个背景下，人们引入了网格的概念。网格之父 Ian Foster 在文献[3]中给出的描述是：计算网格是一个能够为人们提供可靠的、一致的、普适并且廉价的高端计算能力的软、硬件平台。

从广义上来讲，网格就是一个集成的计算与资源环境，或者说是一个计算资源池，它能够充分吸纳各种不同类型的计算资源，将它们转化为一种可靠的、易得的和标准的计算能力。在这种广义的定义下，网格中的资源包括了各种类型的计算机、网络通信能力、数据资料、仪器设备等。实际上，广义的网格概念就是我们平常所说的网格计算(Grid Computing)，Ian Foster 等在文献[4]中给出的描述是：网格的概念就是在动态、多机构虚拟组织之间的协调的资源共享和问题解决。

网格技术的发展经历了如图 1-1 所示的几个主要阶段^[5]。

20 世纪 80 年代中后期，为了满足科研活动对新的高性能计算技术的需要，网格计算的前身——元计算^[2]开始受到人们的关注。元计算又被称为网络虚拟超级计算机。著名的 Globus 项目在 1995 年启动时最初的目的也是提供一个元计算工具包^[6]。SETI^[7]和 Conder^[8]都是这一时期的代表性研究成果。



图 1-1 网格技术发展过程

到 20 世纪 90 年代中期,元计算的概念逐渐被从“电力网格”中借鉴来的“计算网格”的概念取代。然而,这一时期的网格技术研究项目大多是各自为政,没有一个统一的指导思想和规范。Globus Toolkit 较早的两个主要版本 1.0 和 2.0 分别在 1998 年和 2001 年发布。它们都基于 C 语言来实现,其中 GT 2.0 的影响比较大,被 LHC Grid 和 Tera Grid 等著名网格采用。

2001 年,与网络多层体系结构类似的沙漏型多层体系结构被引入网格中。同年,GGF 成立。GGF 及其继承者——开放网格论坛(Open Grid Forum, OGF)^[9]随后积极推动网格技术标准和规范的建立。2002 年,GGF 会议上提出 OGSA^[10]和开放网络服务基础设施(open grid services infrastructure, OGSi)^[11]。这两个规范的提出使得网格技术正式走入面向服务的时代。OGSA 的提出统一了网格系统的体系结构,规范了网格的重要功能和关键组件。OGSi 提出了有状态网格服务的概念。面向服务思想的引入简化了网格资源之间的访问界面和协议,方便了资源之间的互操作^[12]。Globus Toolkit 3 遵循 OGSA 体系架构,并实现了 OGSi 中定义的网格服务标准。

然而,由于 Web 服务技术具有更大的影响力及在工业界更高的支持度,OASIS^[13]提出的 Web 服务资源框架(web services resource framework, WSRF)^[14]在 2004 年后逐渐代替了 OGSi。WSRF 的提出标志着网格技术与 Web 服务技术的最终结合。新的 Globus Toolkit 4 即实现了基于 WSRF 标准的网格服务。

1.3 网格的特点

一般来说,网格解决的问题有三个特征。

(1)资源的异构性。网格中可以用来共享的资源有着极其广泛的类型,包括网络资源、计算资源、存储资源、数据资源等。同一种资源也往往有着不同的表现形式和实现方式,如作业系统的 PBS (Portable Batch System)和 LSF (Load Sharing Facility),传输协议中的文件传输协议(file transfer protocol, FTP)和超文本传输协议(hyper text transfer protocol, HTTP)等。类别不同的资源通过网格进行互联,解决了它们之间的通信和互操作问题。