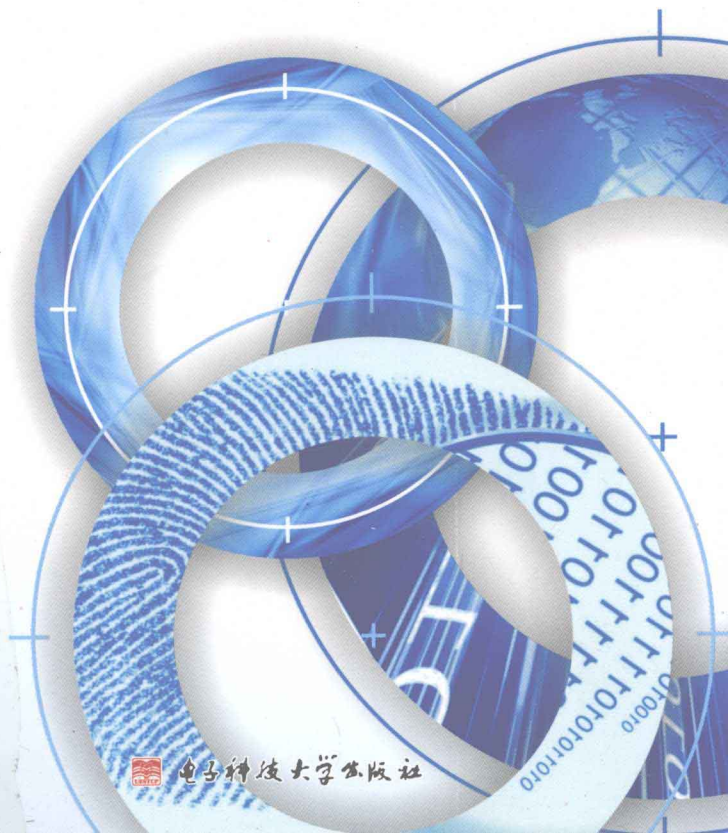


JIYU SHUZIYOUJITI
SHUJUKU XITONG DE
SOUSUO DINGWEI YANJIU

基于数字有机体数据库系统的 搜索定位研究

陈建英 著



电子科技大学出版社

基于数字有机体数据库系统的

搜索定位研究

陈建英 著



电子科技大学出版社

图书在版编目 (CIP) 数据

基于数字有机体数据库系统的搜索定位研究 / 陈建英著.
—成都: 电子科技大学出版社, 2011. 5

ISBN 978-7-5647-0887-0

I. ①基… II. ①陈… III. ①数据库系统—数据检索
IV. ①TP311.13

中国版本图书馆 CIP 数据核字 (2011) 第 137198 号

基于数字有机体数据库系统的搜索定位研究

陈建英 著

出 版: 电子科技大学出版社 (成都市一环路东一段 159 号电子信息
产业大厦 邮编: 610051)

策划编辑: 郭蜀燕

责任编辑: 李述娜

主 页: www.uestcp.com.cn

电子邮箱: uestcp@uestcp.com.cn

发 行: 新华书店经销

印 刷: 四川经纬印务有限公司

成品尺寸: 130mm×185mm 印张 6.375 字数 148 千字

版 次: 2011 年 5 月第一版

印 次: 2011 年 5 月第一次印刷

书 号: ISBN 978-7-5647-0887-0

定 价: 15.00 元

■ 版权所有 侵权必究 ■

◆ 本社发行部电话: 028-83202463; 本社邮购电话: 028-83208003。

◆ 本书如有缺页、破损、装订错误, 请寄回印刷厂调换。

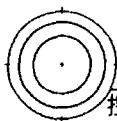
摘 要

随着 IPTV、EC、网络服务 (Web Service)、音乐和电影下载等大规模网络应用的不断发展, 网格、P2P (Peer to Peer) 和云计算等分布式计算技术不断推陈出新。这些技术往往基于现有基础软件平台、通过摒弃传统分布式系统中的某些约束并采用新技术或改进原有技术来解决分布式网络环境下的特定应用问题。

为了寻求面向大规模网络应用的通用计算模式和整体解决方案, 8010 研究室把大规模分布式计算技术推进到基础软件层, 经过多年多个阶段研发出包括 OS 和 DBMS 两个核心软件的“数字有机体系统”基础软件平台并取得了应用的成功。该平台除了必须具备基础软件全部功能外, 还面临着与其他大规模分布式应用系统一样的高性能网络服务问题, 而这一问题的解决在很大程度上依赖于系统内部共享资源的搜索定位机制。

本书对基于数字有机体数据库系统的数字有机体系统搜索定位机制进行了全面研究, 主要工作和贡献如下:

1. 在分析基于网格计算、对等计算和云计算等分布式计算技术的大规模应用系统的搜索定位技术、研究现状和发展趋势的基础上, 本书介绍了数字有机体系统的由来及其研究的目标和意义, 首次把数字有机体数据库系统的研发历程总结为分布式、分布式并行和数字有机体三个阶段并对各阶段成品系



统在网络结构、软件体系结构和应用情况等方面进行了详细介绍。

2. 构建了数字有机体数据库系统基于区/站/节点三级分层和区间对等的混合覆盖网结构,并基于该结构建立了站内分布式并行、区内回溯和区间对等的搜索定位模型,介绍了基于数字有机体数据库系统的数字有机体系统搜索定位的目标、类型、步骤和特点。

3. 提出并构造了包括描述热点资源缓存的缓存信息表、记录搜索定位历史的信息库、描述搜索定位相关知识信息的知识库、提供按资源名或关键字定位的共享资源索引信息表等多种辅助快速搜索定位的特殊数据结构并给出了相应的搜索定位算法。

4. 研究了基于分布式并行技术的数字有机体数据库系统站内搜索定位机制,介绍了执行站内搜索定位的分布式并行事务并对事务执行过程中的服务器节点角色进行了划分,描述和分析了保证站内搜索定位结果正确性和一致性的分布式并行两阶段提交协议,分析了基于分布式并行技术的搜索定位扩展到大规模系统的诸多局限性。

5. 提出了基于位置索引压缩的共享资源索引信息压缩方法,阐述了利用位串表示关键字位置分布信息的原理和方法,介绍和分析了基于位置索引压缩的索引信息发布、汇聚和区内搜索定位算法。实验表明,该压缩方法可达到较高的记录压缩比,对搜索定位效率的提高也起到了一定的作用。

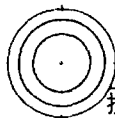
6. 提出了基于多维度 QoS 约束的副本定位方法 MQoSRL,定义并形式化描述了多维度 QoS 约束问题,给出了包括索引分级发布和区间路由等子算法在内的 MQoSRL 全

局副本搜索定位相关算法。分析和应用表明, MQoSRL 具有高效和负载均衡等特性, 同时还可有效降低系统平均响应时间。

关键词: 数字有机体数据库系统; 搜索定位; 混合覆盖网络; 多维度 QoS 约束; 位置索引压缩

目 录

第一章 绪论	1
1.1 课题背景及意义	1
1.2 国内外相关研究综述	3
1.3 论文的主要研究内容及贡献	22
1.4 论文的组织结构和章节安排	23
第二章 数字有机体数据库系统 DODBS 概述	26
2.1 数字有机体系统简介	26
2.2 数字有机体数据库系统研发概述	38
2.3 三个研发阶段的回顾和总结	39
2.4 本章小结	58
第三章 数字有机体数据库系统 DODBS 的搜索定位模型	60
3.1 DODBS 的搜索定位概述	60
3.2 DODBS 的混合覆盖网模型	64
3.3 DODBS 的搜索定位模型	69
3.4 DODBS 搜索定位的特点	80
3.5 性能比较与分析	82
3.6 本章小结	86
第四章 利用特殊数据结构的 DODBS 快速搜索定位	87
4.1 缓存信息表	87
4.2 信息库	92
4.3 知识库	94
4.4 共享资源信息表	97
4.5 实验与分析	102



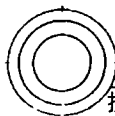
4.6	本章小结	105
第五章	基于分布式并行技术的 DODBS 站内搜索定位	106
5.1	DODBS 站内搜索定位概述	106
5.2	数据目录简介	108
5.3	DPSQL 分布式并行事务及服务器节点角色分析	110
5.4	基于数据目录一致性的分布式并行数据更新协议	114
5.5	数据目录与分布式并行事务的 ACID 特性	124
5.6	站内分布式并行搜索定位算法	127
5.7	分布式并行搜索定位技术的局限性	131
5.8	本章小结	132
第六章	基于位置索引压缩的 DODBS 搜索定位	133
6.1	概述	134
6.2	基于位串的位置索引压缩方法	138
6.3	基于位置索引压缩的信息发布与汇聚	140
6.4	基于位置索引压缩的 DODBS 搜索定位	143
6.5	性能分析与实验	146
6.6	本章小结	153
第七章	基于多维度 QoS 约束的 DODBS 副本定位	155
7.1	现有副本定位分析	156
7.2	MQoSRL 定义及其形式化描述	161
7.3	基于多维度 QoS 约束的副本定位算法	165
7.4	性能分析与实验	170
7.5	本章小结	174
第八章	结论和进一步的工作	175
8.1	研究总结	175
8.2	进一步的工作	177
	参考文献	179

第一章 绪 论

1.1 课题背景及意义

随着计算机网络不断地朝着规模化、智能化、高可靠/高可用和高动态伸缩等方向飞速发展,各种分布式计算技术不断推陈出新,先后出现了以数据、计算和设备等资源整合为目标的网格系统、强调用户间关系平等和充分利用边缘网络资源的P2P系统、以减少用户终端处理负担和提供“按需服务”为目标的云计算系统等具有代表性的大规模分布式计算系统。这些系统基本上都是基于现有基础软件平台,通过采用新技术或改进原有技术来解决大规模分布式网络环境下的特定应用问题,多种计算模式并存的格局说明,到目前为止还没有哪一种系统或技术能够成为分布式计算事实上的标准或提供通用的分布式计算模式。

面对服务需求日趋复杂、性能需求越来越高的大规模网络应用,研究者们试图优化或改造已有系统使之提供更有深度、广度和速度的高性能网络服务,同时,也有把大规模分布式计算技术推进到基础软件层的研究出现,这些研究试图寻求一种既能囊括各种分布式计算技术优点又能满足当前和今后相当长一段时间内大规模网络应用需求的通用计算模式及相关软件。数字有机体系统正是这样一种为解决大规模网络应用问



题、经过多年多个阶段研发并取得了应用成功的基础软件平台,现已获得国家发明专利 1 项、软件产品登记证 3 项、软件著作权 11 项,正在申请国家发明专利 3 项和软件著作权 8 项;得到了包括 IBM 全球技术总裁亚太地区总裁以及朱中梁、李乐民等院士在内的各界权威人士的充分肯定和高度评价。目前,基于数字有机体系统的各种应用系统已在成都市检察机关、电信科技第五研究所等部门和单位正式投入使用,《中国科技成果》2010 年第 14 期一篇名为“赤诚振我中华 领域创新又逢春”的文章报道:“长时间的实际运行证明,系统是稳定的,权威部门的查新表明,系统属于世界首创。”基于数字有机体系统的数字有机体抗毁信息系统于 2010 年 5 月通过了四川省科技厅技术成果鉴定^[1],鉴定结果为“技术上具有重大的创新性……达到了国际领先水平”;相关的学术和技术研究已经大量出现或正在跟进^[2~16]。

数字有机体系统因其自适应、自学习、自免疫、自传播和群体协作等生物有机体特性研发目标而命名,核心软件包括数字有机体操作系统和数字有机体数据库系统。除了必须具备基础软件的全部功能外,数字有机体系统还需解决与其他大规模分布式应用系统(如网格系统、P2P 系统和云计算系统等)一样的两个关键问题——海量数据资源的有效组织和高效利用问题以及高性能的网络服务问题。这两个问题都与共享资源的搜索定位相关,后一问题的解决则在很大程度上依赖于所采用的覆盖网模型及相应的搜索定位机制。

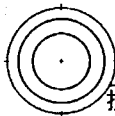
与现有大规模网络应用系统的搜索定位机制相比,数字有机体系统充分利用了自身两个基础软件——操作系统和数据库管理系统——间的紧耦合性,把全局各种类型的共享资源及

其相关信息按照相应的部署策略均匀地分布于系统各相关服务器节点,共享资源相关信息由数字有机体数据库系统以通用的方式管理和维护,系统范围内的搜索和定位全过程则通过调用内部查询访问接口来实现。这种由自己的数据库软件管理和维护共享资源信息的方式可充分利用数据库的查询高效性、事务并发性等优势有效保障数字有机体系统共享资源搜索定位的高效性;以访问接口提供透明搜索定位服务的方式不仅可消除应用程序开发对中间件的依赖性、降低大规模网络应用开发的难度、实现应用程序与共享资源之间的独立性,还可有效保障系统资源访问的安全性并为各种网络应用提供通用的搜索定位服务。

共享资源的搜索定位是所有大规模分布式系统的核心问题之一,不同的计算模式对此有不同的处理方式,但都必然与系统的设计目标、所采用的逻辑网络拓扑结构、资源的放置策略等密切相关。本文对数字有机体系统基于自身数据库进行搜索定位的机制进行全面的剖析,从系统的逻辑拓扑结构设计、搜索定位模型到具体的搜索定位算法以及搜索定位过程中的负载均衡、信息压缩和 QoS 约束等方面进行的深入研究,对降低搜索定位延迟、提高搜索定位结果的准确性和合理性将起到积极的促进作用,同时也是实现高性能网络服务终极目标不可或缺的研究内容。

1.2 国内外相关研究综述

在大规模分布式计算领域,目前广泛流行和采用的技术主



要包括网格计算、对等计算和云计算技术。下面对基于这三种代表性技术的应用系统在搜索定位研究、发展趋势及其存在的主要问题等方面进行综述。

1.2.1 网格计算

1.2.1.1 网格简介

网格 (Grid) 计算是继分布式计算和并行计算后出现的第一种被广泛研究和重视的网络计算模式。被誉为“网格计算之父”的 Ian Foster^[17, 18]将网格定义为:“支持在动态变化的分布式虚拟组织 (Virtual Organizations, VO) 间共享资源、协同解决问题的系统。”换句话说, 网格系统利用互联网把地理上广泛分布的各种资源组织成一个逻辑整体, 就像一台超级计算机一样为用户提供一体化信息和应用服务, 最终实现这个虚拟环境下的资源共享和协同工作。

从定义可以看出, 网格计算的主要目的是共享资源 (包括计算资源、存储资源、带宽资源、软件资源、数据资源、信息资源、知识资源等) 和协同解决大规模网络应用问题 (如: 计算、存储和访问等)^[19], 特别是那些仅仅依靠单一计算设备无法解决的复杂问题。因此, 网格计算侧重利用 Internet 上大量闲置的共享资源, 强调问题分解和分解后的任务在共享资源上执行的管理和控制。其核心任务是问题分解、资源共享和资源组织。

针对问题分解, 网格计算在并行计算理论的基础上提出了针对具体问题进行研究的思路, 因此大多数网格系统都是针对具体类型的应用而研究和建设, 资源共享和资源组织也根据应用的目标而有所不同。但总体来讲, 可把网格分为计算型和数

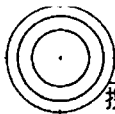
据型两大类, 计算型网格侧重于计算资源的管理, 核心内容是任务管理和任务调度, 数据型网格侧重于存储设备和数据信息的管理和维护, 核心内容则是数据存储和数据访问。

在网络计算领域中, Globus^[20]是最著名的、用于构建计算网格开放体系结构、开放标准的网格开发项目, 它对资源管理、安全、信息服务及数据管理等网格计算的关键理论进行研究并提供基本的机制和接口。Globus Toolkit^[21~23]是该项目开发出来的一套服务和软件库, 用于在各种平台上运行网格计算和网格应用, 帮助规划和组建大型的网格试验平台, 开发适合大型网格系统运行的大型应用程序。在 Globus Toolkit 的基础上, Globus 项目组和 IBM 共同倡议了一个全新的网格标准——开放网格服务体系 (Open Grid Services Architecture, OGSA)^[24,25]。OGSA 由服务体系结构、服务基础设施、服务安全体系结构和数据库访问与集成等几个方面的内容组成。

到目前为止, Globus 已经成功应用于全球数百个站点和几十个主要的网格计算环境, 包括美国国家航空航天局信息网格 (National Aeronautics and Space Administration Information Power Grid, NASA IPG)、欧洲数据网格 (Data Grid)、美国国家技术网格 (NTG) 和地球系统网格 (ESG) 等。Globus Toolkit 具有较为统一的国际标准, 大多数网格项目都基于它提供的协议及服务建设, 大多数学术研究也是基于 Globus 网格平台。

1.2.1.2 网格中的搜索定位研究

在网格开发工具中, Globus 最为广泛流行, 算得上是一种事实上的行业标准。因此, 下面就以它为例说明网格中的搜索定位服务机制。



Globus 的搜索定位是指在网格系统中查找到满足执行作业所需的操作系统类型、版本、处理器类型、内存大小、网络带宽、负载程度以及所装软件等条件的共享资源。

Globus 通过监控和发现系统 MDS (Monitoring and Discovery System, MDS)^[26, 27]实现资源的发现和描述以及服务和计算的监控。MDS 是 Globus Toolkit 的信息服务组件, 由网格资源信息服务 (Grid Resource Information Service, GRIS) 和网格检索信息服务 (Grid Index Information Service, GIIS) 两大部分组成, 包括 Index service、Trigger service 和 Aggregator service 等主要组件。

Globus 实现搜索定位的原理主要基于资源注册和发现机制: GRIS 通过网格资源注册协议 (Grid Resource Registration Protocol, GRRP) 向 GIIS 注册资源; 网格应用程序通过网格资源查询协议 (Grid Resource Inquiry Protocol, GRIP) 向 GIIS 查询虚拟组织 (VO) 中的网格资源或直接向某个资源节点上的 GRIS 查询该节点的资源及其属性; 通过使用轻型目录访问协议 LDAP (Lightweight Directory Access Protocol)^[28]提供统一的界面来描述分散的资源。

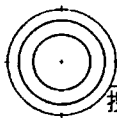
Globus 早期的搜索定位服务基于集中式目录结构, 目前已在欧洲数据网格等多个项目中得到应用^[29], 这种结构虽然实现简单, 但是限制了系统的可扩展性和可靠性; 而后, 欧洲数据网格 (Data Grid) 项目组的研究人员提出了一种层次式目录结构^[30], 这种结构因所有逻辑文件必须在根节点中出现导致了庞大的根节点数量, 限制了数据网格系统的可扩展性, 同时可靠性和查询性能都比较差; Globus 项目组和欧洲 Data Grid 项目组联合提出的 Gigggle^[31]框架则针对数据网格系统提

出了一个通用的层次式副本目录结构,通过结合应用的具体特征设置和改变参数的取值来改变副本目录的结构和性能,实现了系统搜索定位的灵活性,但系统扩展性和可靠性等性能却因此受到影响。

国内外关于 Golbus 资源定位学术研究的文献也比较多,如: Ripeanu 和 Foster 提出的分布式副本定位方法^[32],这种方法采用 Bloom Filter 技术在本地节点上压缩存储系统中全部的副本定位信息,查询性能良好,但存储和更新维护开销较大;李东升等人提出了动态自适应性的副本定位方法 DSRL^[33,34],采用动态均衡映射将全局副本定位信息均衡分布在多个宿主节点上,具有较好的可扩展性、可靠性和性能,但该方法中每个宿主节点都要维护全局副本定位索引信息,存储和更新开销也很大;Tree-based^[35] 副本定位与选择相结合的机制 TRLS 从系统代价和经济代价着手定位系统认为开销最小的副本;Ann L. CHERVENAK 等人则通过建立全冗余分布式副本定位索引信息解决副本定位可靠性和可用性问题^[36]等。

1.2.1.3 发展趋势

网格计算是 20 世纪 90 年代提出的概念,曾经被誉为继 Internet 和 Web 之后的“第三个信息技术浪潮”,在世界上引起了前所未有的关注和重视,许多国家为此投入了大量的资金。从实际应用情况来看,由于受限于具体应用的实现,尤其是具体应用中问题分解和并行算法的实现,网格计算并没有成为普遍应用的计算模式。但是,作为解决大规模复杂问题的有效手段,在促进许多科学问题的解决,例如地球物理、气候分析、寻找外星人和数学问题等方面,仍然具有重大的价值。



总体来看, 网络提出了充分利用 Internet 上大量闲置资源的思想, 并且通过一系列的中间件实现了闲置资源的管理和利用, 对推动网络计算的发展有着重要意义。但是, 学术界和业界对于网络的研究都在逐步向特殊化和应用方向转移: 信息网格、语义网格和知识网格等专用网格成为当前网格研究的热点; 网格在各种行业中的应用也成为新的研究方向, 典型的应用领域有 Web 服务、地理信息系统、数字图书馆等。

根据 Google 的学术趋势统计, 网格计算的研究已经过了关注顶峰期。随着其他计算模式的兴起, 网格计算已不再作为一个单独的研究方向, 而是朝着与对等计算、云计算等计算模式相融合的研究方向发展。

1.2.2 对等计算

1.2.2.1 对等计算简介

对等也称为 P2P (Peer-to-Peer, P2P), 是一种几乎和网格计算同时兴起的在 Internet 上实施网络计算的新模式, 本质上也属于分布式计算的范畴。和网格计算相比, 对等计算的最大特点是强调用户之间相互“平等”和相互共享资源。

对于“P2P”的概念, 存在较多表述各不相同的定义, 如 Intel P2P 工作组^[37]的定义: “通过在系统之间直接交换信息来共享计算机资源和服务。”^[38]Aberdeen 大学的 Alex Weytsel 的定义: “以非客户端的方式利用 Internet 周边的设备。”^[39]IBM 的定义: “P2P 系统由若干互联协作的计算机构成且至少具有如下特征之一: 系统依存于边缘化 (非中央式服务器) 设备的主动协作, 每个成员直接从其他成员而不是从服务器的参与中受益; 系统中的成员同时扮演服务器与客户机的角色; 系统应

用的用户能够意识到彼此的存在，构成一个虚拟或实际的群体。”诸如此类的定义都或多或少提到了对等网络构架和以前客户端/服务器构架（Client/Server，C/S）的不同。不难看出，P2P 把网络计算模式从集中式引向分布式，也就是说，网络应用的核心从中央服务器向网络边缘的终端设备扩散：服务器到服务器、服务器到 PC 机、PC 机到 PC 机、PC 机到 WAP 手机等，所有网络节点上的设备都可以建立 P2P 对话。

P2P 系统的分类方法很多，主流方法主要基于 P2P 系统拓扑结构的耦合度和分散度^[40~42]，耦合度用于衡量 P2P 系统拓扑构造过程可控的程度，分散度用于衡量 P2P 系统的拓扑结构对中央服务器的依赖程度。

通常情况下，P2P 系统的拓扑构造过程可分为严格受控于某种机制和动态、非确定性构造两种情况。因此，P2P 系统可根据耦合度分为如下两大类：

1. 非结构化 P2P 系统

非结构化 P2P 系统的节点之间逻辑拓扑关系通常较为松散并且具有较大的随机性，资源或资源的元数据信息一般只放置在本本地，放置策略通常与 P2P 系统的拓扑结构无关。

非结构化拓扑结构的实现和维护相对简单，可支持灵活的资源搜索条件，但高效的资源搜索通常较为困难（通常都是采用广播搜索、随机转发和选择性转发等方法），因此，这种 P2P 系统通常适用于由大量自治性强的节点组成、对服务质量没有严格要求的应用，如 P2P 文件共享应用等系统。

非结构化 P2P 系统还可进一步根据分散度分为三类：集中式非结构化 P2P 系统（如：Napster^[43]和 BitTorrent^[44]等）、全分布式非结构化 P2P 系统（如：Gnutella^[45]和 Freenet^[46]等）