

智能型分布式信息系统的 结构与功能

陈剑 (武汉中南财经政法大学信息学院信息系 430064)

摘要: 本文描述了在网络环境中分布信息系统的结构, 剖析了分布式信息系统的功能, 重点阐述了联机分析与知识处理技术在分布式信息系统中的作用。同时也讨论了分布式信息系统的设计原则和实现技术。

关键词: 分布式信息系统 多层系统 分布式数据仓库 知识库 知识处理

1 引言

计算机技术特别是软件技术的发展使得分布式的进行工作成为今天广泛流行的工作模式, C/S 模式就是其中一例。然而, C/S 模式中, 前端处理机只支持联机处理的功能, 距离人们的期望还很远。同时, 管理者对信息系统的辅助决策能力的要求也日益提高, 这就要求信息系统在可以进行信息存储、归类的基础上, 又要有部分分析、发现隐藏信息和辅助决策的功能。人工智能与知识工程的发展使得这方面的应用开始普及。现代, 从事数据分析与处理的研究者已经把人工智能的一些机制用于数据挖掘, 进而支持基于知识的处理。所以, 无论从技术角度, 还是从商业、贸易发展的角度看, 分布式信息系统是信息系统发展的必然趋势。对比于传统系统的一些缺点和不足之处, 它有极为强劲的优势。

2 C/S(客户机/服务器)模式的缺陷

在现代广泛流行的 C/S 模式(即客户机/服务器模式)的条件下, 信息

的加工与处理存在以下的不足之处:

(1) 当需要交换的信息量增大时, 用于后台支撑的数据库系统被设计的更加复杂;

(2) 数据库如果出现故障乃至崩溃, 将严重影响各个终端的信息处理能力;

(3) 服务器的压力会随着访问量的增大而增大, 极大的影响与客户的信息交互;

(4) 不提供个性化的服务界面;

(5) 缺乏智能处理的能力, 在辅助决策时较为机械;

这些问题来源于 C/S 模式系统结构的组织方式。在 C/S 模式中, 客户机只是机械的充当配角, 而将大部分工作交由服务器来完成, 当业务量急剧上升时, 服务器就会难堪重负。实践证明, 网络环境下, 当工作量增大时, 往往在客户端会得到服务器忙、出错、联结无效等信息, 但试图再次访问时, 很有可能轻易进行登陆, 这就是服务器处理问题能力的具体表现。另外, 许多商业网站虽然保留用户的部分个人信息, 却不给用户提供

个性化的服务界面, 对任何一个用户一般是用同一副面孔, 这与进行个性化服务的服务发展趋势是矛盾的。对于商业公司来讲, 其业务数据库如果崩溃, 后果可能是灾难性的。举个例子来

说, 现代银行的客户信息大多都存放于数据库中, 对于一个每日交易量上亿元的银行, 数据库哪怕是出现微小故障都有可能巨大的损失。所以, 数据库的维护工作就显得特别重要。

为了有效的解决上述问题, 必须改变以 C/S 模式为基础的信息系统, 将计算与处理任务合理分配。同时, 进一步强调信息系统的整体性、自治性和智能性。为了达到这个目的, 必须将信息系统构造成一个跨地区的综合平台, 严格的定义计算机之间的并行机制, 数据库也必须认真进行分片和引入合理的联结机制和数据保护措施。为了有效进行信息管理, 符合网络特性进行工作, 充分利用网络的特点和优势, 导入分布式信息系统是必要的。

3 分布式信息系统的系统结构

所谓分布式信息系统,是指以计算机网络为基础,将系统的数据与功能分布在地理上不同位置,通过自然的数据与功能联结而进行辅助决策的信息系统。严格来讲,分布式信息系统最为适合分布式组织的企业和单位。

作功能的能力;

(4) 坚强性与优美降级功能。当系统出现故障时,系统可以降级为低级的系统,但系统本身的降级并不损失系统功能的完整性,待故障部分修复后再自行升级至原状态;

(5) 自动生成的能力。利用软件自动化的技术来生成系统的一些部分,用以支持用户的工作,也应是分布式操作系统的功能之一;

(6) 文件的分布式管理和透明操作。即为用户提供逻辑文件服

合,用来支持管理人员的决策。数据仓库是一个动态变化的动态变化的系统,数据就像“物品”一样被存储,整合,理解为信息。换言之,对于决策者来讲,数据仓库是一个信息仓库。数据仓库与传统数据库的最大区别是前者拥有数据分析的能力,而数据库系统则不具备。对于基于网络环境、在地理上分布的数据仓库,称之为分布式数据仓库。

分布式数据仓库的系统结构如图2所示,其中每个结点的局部结构如图3所示:

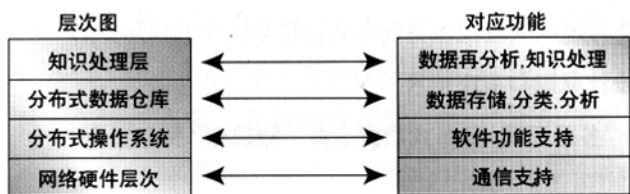


图1 分布信息系统的层次

由于分布式信息系统是以计算机网络为基础而构建的,所以它必然是一种分层结构。

分布信息系统的层次性描述如图1所示:

对于网络硬件层次,一般可以利用TCP/IP协议进行通信。这方面已经有了许多成熟的技术用以保证数据的安全传输,本文不加详细讨论。

3.1 网络操作系统层

任何软件必须获得操作系统的支持,分布式信息系统也不例外。支撑整个系统必须有健壮并且功能强大的分布式操作系统(Distribute Operating System)。一个分布式操作系统至少应具有以下的功能:

(1) 移动计算的能力与高度的并行机制。运算能力的强大是分布式操作系统的优势;

(2) 合理协调系统内部的并发行为。特别是异地计算机之间的同步工作能力;

(3) 智能、自治的对象机制。分布式操作系统应该合理地规划对象的功能,对象应具有携带数据和拥有操

务,隐蔽文件的物理特性。

3.2 用于支撑数据管理的分布式数据仓库

传统意义上的信息系统是由数据库系统进行支撑的。数据库技术经过数十年的长足发展,已经取得了辉煌的成就。但一个地理上分布的企业往往拥有许多数据库,这些数据库的设计与组织是针对各个部门而进行的。对于统一业务的目的来讲,数据在这些异构的,孤立的数据库上进行交换是不合理的,所以必须有效统一多数据库为一个整体。由于这些数据库本身是分布的,所以对于整合后形成的数据库系统也应该是分布的。

然而,企业对数据的要求已经不再只是简单的处理,而是要在数据中发掘潜在的经济因素。数据作为一种资源,需要用一些统计与决策分析的方法加以分析处理,这就要求在数据库分布的技术基础上进行整合与加工,从而产生了数据仓库技术。严格说,数据仓库是一个面向问题的集成的不可更新的且随时间不断变化的数据集

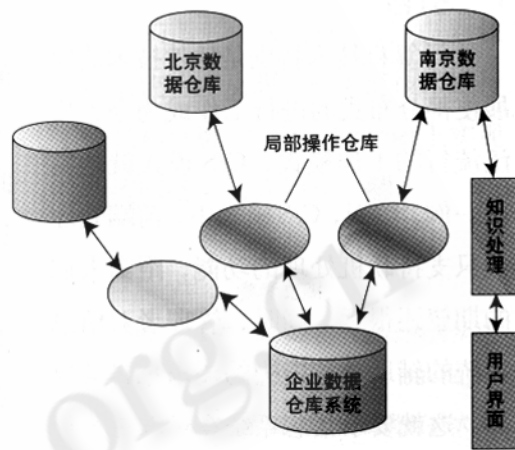


图2 分布式数据仓库平面图

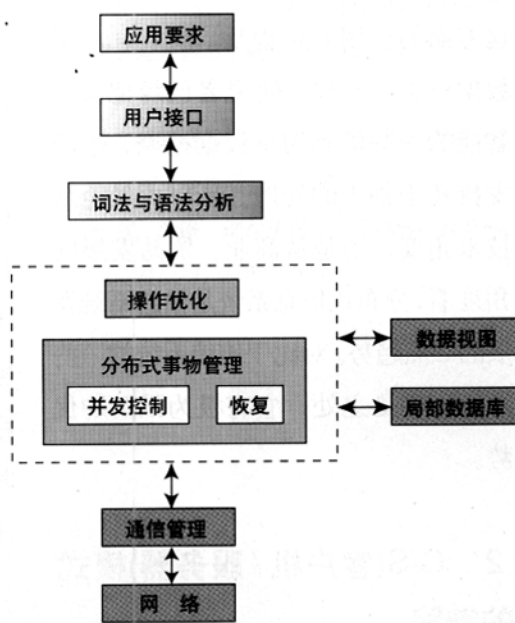


图3 局部数据仓库的结构

3.3 基于 Web 的信息抽取

数据仓库在应用于企业内部,它的支撑数据管理的能力是足够的。但对于企业外部信息处理却遇到了困难。因为企业内部的信息是有限的数据集,而企业外部的信息是无限的,这一点在基于Web的查询上表现的淋漓尽致。Web上拥有大量分布的动态的信息,支持企业信息系统的数据库面对浩瀚如海的Web信息是无能为力的。然而,忽视Web上大量有价值而且有免费的信息又是巨大的损失。但对于Web信息,信息系统应利用不同的方法来发掘数据和抽取相关信息。由于这是一个全新的课题,面对的是近乎无限的、不可遍历的数据,所以在方法上仍不是很成熟。下面给出一些广泛使用的应用技术和一些成熟的方法。

(1) 搜索引擎:许多搜索引擎提供了对于相关内容的查询。包括使用机器人进行爬行,使用层次树的链接和基于Agent的检索系统;

(2) 数据挖掘:利用搜索引擎的Web页一般是粗糙的,必须进行进一步的深度发掘。一般使用的方法有:联系分析,预测分析,聚类分析等;

(3) 日志维护:对于经常抽到的信息源,应维护日志队列,下次检索时可以直接抽取。

网络条件下,信息系统处理的信息不仅包含企业内部信息,还应包括外部信息,例如客户信息、消费者反馈信息等。能够合理的处理外部信息也是分布式信息系统的功能之一。

3.4 分布式信息系统的数据分析与知识处理能力

构造信息系统的目的在于合理地管理、分析、处理信息,从中发现潜在的商业价值。然而,孤立的数据一般很难反映信息的本质,信息必须在

一系列连续并具有相互联系的数据中得到体现。能够主动地去发现信息,把管理信息系统推向决策支持系统的高级阶段,正是分布式信息系统较之于传统的集中式信息系统的优势所在。现代的管理正日益向知识化、科学化、自动化转化。同时,以人为本的管理模式也要求信息处理过程更知识化与自然

化。在知识时代,信息系统的重要支撑应是知识。计算机处理知识的能力正在日益加强。一旦人工智能与知识工程知识技术的成就与信息系统完美地结合起来,信息系统辅助决策、自主选择信息的能力就会大大加强。对于从事信息选择的工作人员来讲,把一些机械、重复的工作交给计算机来做,工作效率无疑是一次质的飞跃。

数据仓库分析与处理数据,Web信息的抽取都必须经过分布式信息系统的选择与过滤,加工与析取,因此增强信息系统的知识处理能力是一个必须解决的问题。知识处理的能力必须高度依赖于已有的知识。真正意义上的智能系统必须具备智能化的能力。只有计算机的信息处理的表现形式更加智能化、自然化地贴近于人,人性化地与人进行交互,才能体现存在于计算机的高度智能。现实条件下,智能推理技术正逐渐向这方面发展,专家系统业已问世,知识库的存取已经有了成型的方法,支撑知识处理的技术正日渐成熟起来。

下面讨论分布式信息系统的知识处理能力,从而体现在网络环境中智能技术的应用。

4 分布式信息系统的知识构造

4.1 知识处理的框架结构

知识处理的基础是知识和进行基于知识处理的工具。在知识处理中,知识库居于核心地位。一个典型的知识处理系统的组织如图4所示,详释如下:

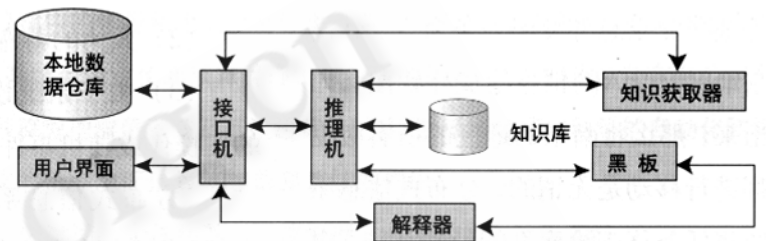


图4 知识处理系统的组织图

知识库是存放专门领域知识,包括事实信息,规则等。知识库应该是开放的,它允许对于规则的加入和删除。凡添加和知识必须具备一致性原则,即加入的知识不允许是矛盾的;

推理机实质是个程序库,它用来执行检索和进行推理;

接口机用于和数据仓库与用户界面联系,它将数据仓库中的数据进行抽出,同时负责转化外部的信息形式为推理机可以识别的状态。对于用户,接口机可以提供人机交互功能和自然语言的识别;

黑板实质上是个中间数据库,它用于存放推理过程的中间结果。同时,它也是个动态的知识库,控制着推理进程。

知识获取器也称学习器,机器学习的能力由它来体现,对于一些模糊、经验的规律可以进行抽取并引入知识库。

解释器用于表达系统的推理过程与推导的轨迹。

4.2 智能辅助决策能力

在分布信息系统的各个层次中,网络与分布式操作系统提供了数据传输与计算任务的并发执行;分布式的数据仓库为系统进行了必要的管理、备份、维护和一致性检查,并能抽取数据进行简单的分析;而高级的管理模式--决策支持系统的功能则由知识处理系统来完成。例如,某公司总裁在北京总部发出一条命令,将南京地区产品A的销量绘成柱状图打印出来,那么将南京的地区的所有数据都进行移动是无谓的。分布式信息系统将任务移交给南京的处理系统,由南京的处理系统向南京的数据仓库进行查询。处理结果会交给系统进行回传。但绘制柱状图的工作应交由北京的处理系统来进行。如果在南京绘图回传,显然信息的传递是冗余了。

辅助决策的功能必须严格区分决策目标所关心的对象的行为。信息系统可以根据当前的决策目标反映相关联的数据。对于关联数据,基于规则的抽取是完全有效的。例如IBM公司是著名的巨型机生产商,那么了解巨型机的信息时,根据“巨型机→IBM”规则,同时进行抽取IBM公司的信息反馈给决策者是合理的。当然,这条规则是经验规则。

知识规则也是推理和决策的依据。例如,“石油→化工产品”就是一条知识规则,因为石油是生产大多数化工产品的主要原料。石油价格的波动无疑会影响化工产品的价格。同样,“石油→汽油”也是知识规则。诸多规则在一起往往引起链式反应。例如,“石油→汽油”“汽油→汽车”,“汽车--拥有汽车者”。那么“石油→拥有汽车者”,这种条件的建立就是当石油涨价时,汽车拥有者花费在燃料上

的钱就会增多。依据这种机制,系统就可以模糊的辅助决策。

5 分布式信息系统的健壮性与坚强性

系统的安全是系统设计时必须考虑的,系统必须能够抵制三类破坏:

- (1) 硬件损坏的自然破坏;
- (2) 人为的蓄意破坏;
- (3) 由于系统设计不当而导致

的误操作或随机事件的破坏;

分布式信息系统由于层次繁多,所以抵御来自个层次的破坏是相当重要的。常用的手段有:(1) 访问权限控制;(2) 加密保护;(3) 数据备份;(4) 优美降级。

权限访问控制在系统中是极为必要的,例如全国软件专业考试可以在网上查询分数。但不允许考生修改分数,这就是权限控制的能力。对一个公司中,每个层次的管理人员不能越级访问和修改公司数据,这都是对于分布式信息系统的要求。同时,对于黑客、病毒、分布式信息系统同样要具备抵抗能力。一方面,信息系统要从Internet上抽取大量数据,另一方面,信息系统要严格检查数据(包括Web页面、程序等)。万一遇到了不可恢复的破坏性事件或操作,系统通过使用备份可以回到初始状态。数据的安全与一致性问题仍是个研究课题,它在极大意义上影响着系统的健壮性。同时,系统中有一个或几个计算机或数据通路发生故障时,其余部分可以自动重构成一个新的系统,该系统可以接替原系统的工作并且在极大程度上继承原系统的功能,直至故障排除,这就是分布式信息系统的坚强性。

6 智能型分布式信息系统鸟瞰

上文逐层剖析了分布式系统的各个层次。为了给读者留下一个整体印象,笔者将功能与层次对应了起来。为了更加明晰分布式信息系统的结构,下面自下而上列举了系统的功能与层次:

网络硬件层:由异构的分布式计算机系统构成,完成数据的传输功能,向分布式操作系统提供服务;

分布式操作系统层:支持移动计算,对象迁移,向上层软件(数据仓库系统)提供透明服务,屏蔽物理细节,管理和组织下层硬件和系统文件;

分布式数据仓库系统:提供数据的组织、存储、抽取、移用、保护功能,进行简单的联机分析,提供知识处理所必须的数据,它是分布式信息系统的有效后台支撑;

知识处理层:对数据仓库中的数据进行分析、处理,为提供辅助决策功能,提供智能化的人机接口,进行自然语言与计算机可识别形式的转化;

管理决策层:决策者根据反馈的决策信息进行高级决策,依赖人的高级智能。■

参考文献

- 1 Lawrence S, Giles C L. Accessibility and distribution of information on the Web. *Nature*, 1999, 400, 107-109.
- 2 Haag. s等著,《Management Information System for the Information Age (Second Edition)》,严建援等译,北京机械工业出版社,2000.6.
- 3 W. H. Tumon 《Building the Data Warehouse (Second Edition)》,北京机械工业出版社,2000.
- 4 王克宏等著,《知识工程与知识处理系统》,北京清华大学出版社,1994.
- 5 徐洁磐等著,《知识库》,北京科学出版社,1999.