

文章编号 1000 - 5269(2008)01 - 0084 - 04

分布式 workflow 技术研究综述

张 宁¹, 余 霏²

(1. 贵州财经学院 信息学院, 贵州 贵阳 550004; 2. 贵州高速公路开发总公司, 贵州 贵阳 550001)

摘 要: 分布式 workflow 管理系统是当前 workflow 技术研究的趋势和热点。对分布式 workflow 系统的研究现状进行综述, 主要包括从理论和实现两个层次深入探讨了分布式 workflow 系统的解决策略和方法, 并对一些具体的实现机制进行了讨论和比较, 介绍了一些成熟的分布式 workflow 系统, 分析了支持分布式 workflow 的系统体系结构和系统中相关的分布式实现, 并且指出了理论研究的不足和系统实践的未来趋势和发展方向。

关键词: 体系结构; 分布式实现; workflow 管理系统

中图分类号: TP311 **文献标识码:** A

The Summary of Distributed Workflow Technology

ZHANG Ning¹, YU Fei²

(1. School of Information, Financial and Economics of Guizhou College, Guiyang 550004;

2. The development of Guizhou Expressway Company, Guiyang 550001)

Abstract: Nowadays, there is a tendency that research more focus on the distributed workflow management system. From the theory and implementation two levels, deeply discussed the solution strategy and the method of distributed workflow system, and also carried on the discussion and the comparison to some concrete realization mechanism, introduced some mature distributional workflow systems, analyzed the system architecture which supported the distributional workflow realization, pointed out the fundamental research insufficiency and the system future tendency and the development direction

Key words: System Architecture; Distributed Implementation; Workflow Management System

0 引言

当前, workflow 技术成为实现企业业务过程自动化与业务过程重组的核心技术。从计算机的角度上看, workflow 经营过程的一种计算机化的表示模型, 定义了完成整个过程所需要的各种参数, 这些参数包括对过程中每一个步骤的定义、步骤间的执行顺序、条件、以及数据流的建立、每一步骤由谁负责以及每个活动所需要的应用程序^[1]。

workflow 管理系统的体系结构可以是集中式, 也可以是分布式。早期大多数的研究范围都过于狭窄, 而且并不强调商业过程。近来, 许多高一层次的设计都提出了商业过程的管理模型^{[2][3][4][5]}, 但其中大多数都是集中控制。在集中式 workflow 引擎的控制结构下, 是由一个 workflow 引擎来控制所有计算机上活动的执行, 这种集中式的 workflow 引擎处理方式在系统的可靠性、可扩展性、实用性以及吞吐量等方面都不能满足企业执行大规模复杂应用的需求^[2]。随着网络技术的发展、业务过程本身的分布性以及信息资源的松散耦合, workflow 必然要以分布的形式出现。

* 收稿日期: 2007 - 11 - 11

基金项目: 贵州省科技厅 07 年重点攻关项目 (黔科合 GY 字 (2007) 3021)

作者简介: 张 宁 (1981 -), 女, 硕士, 讲师, 主要研究方向: workflow, 计算机集成与协同;

余 霏 (1972 -), 男, 硕士, 工程师, 主要研究方向: 软件工程, 计算机语言。

当前, 分布式的工作流系统成为研究与应用的热点, 本文将从底层通讯基础结构、系统协作方式等方面探讨分布式工作流管理系统的理论基础和技术实现。

1 分布式工作流系统的实现技术

系统所采用的底层通讯基础结构将关系到系统中的各个组成部分之间以什么方式来进行连接, 这个基础结构是分布式应用赖以生存的基础^[2]。

1.1 COBRA 技术

CORBA 是开发大型、复杂的分布式应用的首选标准。CORBA (分布对象请求代理体系结构) 是由对象管理组织 OMG 制定的一个工业标准, CORBA 为不同的操作系统、语言、网络协议和硬件结构间提供了应用层的互操作性, 实现了从用户到用户的操作模式。因此, 能够轻松实现异域产品的合作开发的 CORBA 规范, 大大简化了分布式对象的事件触发机制, 为分布式的工作流管理系统提供了可靠又可行的底层通讯机制, 使得远程透明的互操作成为可能, 也提高了应用开发的效率^[7]。例如开源工作流 SHARK 引擎的产品实现, 就是基于 W3MC 标准, 严格执行 W3MC 和 OMG 规范, 通过它的 POJO (纯 Java 对象) 接口, Shark 提供一个 CORBA 接口, 这样它的 COBBA 客户端就可以和部署为 CORBA 服务的 Shark 来通讯, 而且 Shark 的 CORBA 的实现方式很完善, 通过它, Shark 可以成为一个很快运行的系统。

1.2 移动代理技术

文献 [8] 认为, 代理 (Agent), 可看作是许多在异构的计算机网络的结点中移动的对象, 通过其自身控制, 能够使用结点资源完成任务。文献^[9]认为代理是指位于某些环境的计算机系统, 为了符合其系统设计的目标, 代理在其所处的环境下具有自主的能力, 不受人为或其他代理人的干扰, 并能掌控自己的行为以及内部的状态。以代理为基础的系统, 本质上是主动的、自主的以及模块化的^[10]。所谓主动的, 意思指代理人能主动处理工作流程的作业交换, 如侦测事件发生, 作业流程跳转等; 代理人自主的特性则说明其能主动处理任务, 不受干扰; 模块化代表代理人能根据系统的需求进行专业分工。

所以, 利用可移动代理的持久能力, 可以间接保证系统的可靠性, 并且这种方式特别适合复杂的计算环境, 如支持移动计算等。

1.3 分布事务处理技术^[6]

事务工作流是由 Georgia 大学 Amit Sheh 等人最先提出的, 这一概念强调了与工作流密切相关的事务属性。一个事务工作流包含了多个任务的协作运行, 这些任务可能要访问到异构的、自治的、分布的数据库系统, 任务间的协调是通过基于相互依赖的控制流方法进行描述的, 它为每一个任务定义了执行的先决条件。虽然在系统的可靠性与容错恢复上表现出色, 但目前事务工作流还处于初级阶段。

2 工作流引擎的协作过程设计

工作流引擎是工作流管理系统的核心模块, 其分布协作实现了更高层意义上的分布, 文献^[11]认为工作流管理系统的用户通过工作流管理系统提供的用户界面 (例如工作流客户端或 Web 浏览器) 发出工作流执行服务申请, 然后由多个工作流引擎并行协作的方式完成工作流执行服务, 而并行是通过将工作流执行服务分解为若干子任务, 由特定的工作流引擎完成特定的子任务而实现的。文献^[7]提出了一种新的基于过程模型的多工作流引擎协作, 即通过总控工作流引擎动态分配给各分引擎工作流的任务, 由分布工作流引擎根据连接弧信息自动进行互操作, 多工作流引擎协作完成工作流, 无须实时向总控工作流引擎注册, 真正体现了分布的内涵, 解决了总控工作流引擎的运行瓶颈。

文献 [6] 认为, 目前有两种方法可以实现分布环境下的应用协作。第一种方法是在分布式工作流系统的运行过程当中将有关被执行过程的所有信息 (包括模型定义和实例运行状态) 都打成一个包, 并在整个系统内传递这个包。这种方案的缺点主要是包含过程所有信息的包将非常庞大, 加上为了实现活动的并行性需要在系统内将其复制多份, 因此会对系统网络造成非常大的负担, 从而降低了系统运行的性能。第二种解决方案是对模型进行编译, 事先确定执行的地点, 在运行过程中, 需要传递的只是实例的状态和以前计算的结果, 从而避免了由于大量数据传递而造成的数据库和服务器的过重负担。

3 相关分布式工作流系统体系结构介绍

目前, 工作流产品或研究团体开发的原型系统种类非常多。下面介绍几个发展时间较长、产品相对

成熟、影响比较广泛、具有一定代表性的 workflow 系统的体系结构。

3.1 Mentor

Scaarland 大学的 Mentor (Middleware for Enterprise-wide Workflow Management) 项目定义了一个可跟踪的工作流的架构, 通过使用 TP 监控器来实现容错, 并使用 CORBA 来用于异构商业组件间的通讯和集成支持。与其他的体系架构类似, 当活动处于开始和终止状态时, 通过任务表管理器来交换数据。其初始的原型系统稍具局限性, 今后的主要工作是对业务流程动态变化的支持, 以及非完整工作流的事务回滚机制的研究^[12]。

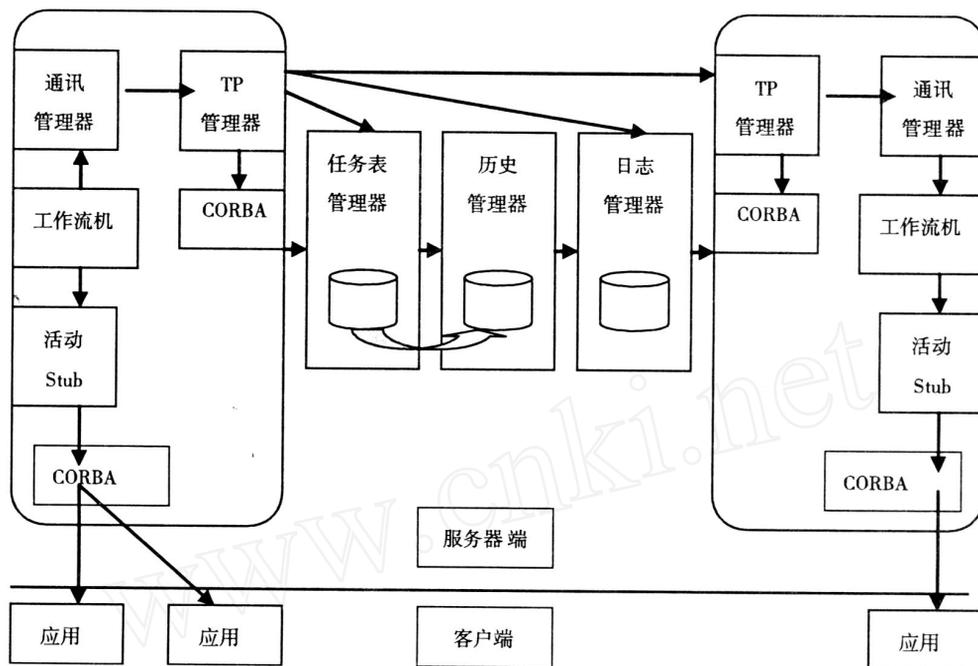


图 1 Mentor 的系统体系结构图^[6]

3.2 Exotica / EMQM

Exotica / EMQM 的体系架构有两个具体的目标, 其一是研究设计完全分布式 workflow 产品的效率, 其二是要分析通过持久消息机制来取代集中数据库后带来的柔性体现^[13]。

它是基于原有的 BM 的工作流管理系统 FlowMark, 通过扩展其功能而实现的一个能够支持大规模复杂应用的高性能分布式 workflow 管理系统, 整个 workflow 管理系统由许多具有自治能力的节点组成, 节点之间通过持久消息 (Persistent Message) 的方式来实现信息传递, 即通过可靠的消息队列在节点之间传递相关的活动信息, 避免了在过程运行之中不断与服务器通信的瓶颈。而且如果一个节点由于故障停止运行, 其余节点还能继续工作, 极大提高了系统的稳定性、可扩展性和柔性。

3.3 INCAS

Barbara^[14]等人提出了分布式 workflow 管理的 NCA s (Information Carriers), 在这种模型中, 每一个过程的执行都与一个信息载体有关, 这个信息载体可看作是包含了执行过程必要的信息对象。这个基于 NCA s 的工作流模型, 用于支持动态 workflow 业务流程的执行, 其流程在各个自治的执行单元里执行。在该体系架构里, 业务流程的定义和 workflow 数据都被封装在一个称作 NCA 的容器里^[15]。

3.4 WONDER^[16]

WONDER 是用于执行大规模范围内 workflow 的分布式体系架构。该架构基于移动代理技术, 即 workflow 的活动可以跨越不同的用户工作站得以执行, 其 workflow 活动的移动按照经代理解释后的 workflow 计划进行。这个特征提供了分布式的控制, 并且定义了一系列的附属组件用于支持代理的移动和 WfMS 的需求, 因而这样一个分布式控制和数据定义方式, 以及大规模范围内 workflow 的执行和管理, 为其应用提供了容错机制。与 NCA 模型类似, WONDER 并非是定义容错的补救规则, 而是通过检查点规则, 来将代理的副本放置在系统的主机上, 并增加了诸如角色的编辑、管理和动态分配等 NCA 所不具备的功能。

目前有许多由“面向消息的中间件协会，MOMA (Message Oriented Middleware Association)”发起的标准化工作，产生了这样一些产品：DEC 的 MessageQ, Transarc 的 Recoverable Queuing Services, Novell 的 Tuxedo/Q, 和 BM 的 MQSeries, 这些产品的特点都是十分适合高度分布的应用。

4 现有分布式 workflow 管理系统的不足

以上提到的项目都是在工作流管理系统中使用了移动代理技术和 CORBA 规范，这些研究主要是通过技术的应用，为不同组织提供可互操作并集成的商业过程。这些研究同时强调了当遇到动态业务过程变化或实例执行失败时，动态可配置 workflow 模型的优势所在。然而，这些项目却缺少对支持大规模 workflow 需求的柔性技术的研究。

workflow 引擎的分布虽然大大解决了集中式控制造成的通讯瓶颈，但一定程度上导致系统的复杂性增加，数据一致性维护相对困难。当前的相关研究中，过分强调 workflow 任务的逻辑性研究，忽略了 workflow 任务的动态性和重组技术研究，即现有的 workflow 模型描述了 workflow 的逻辑、时间和性能三个静态层次，缺乏动态层次的描述，同时忽略了 workflow 本身效率、有效性、可行性等方面评价方法的研究，也缺乏按照任务优先等级不同的原则动态优化配置 workflow 的有效方法。

5 结束语

当前，从 workflow 的发展过程和趋势上看，分布式和柔性技术倍受关注，相关的体系结构、模型定义和应用标准成为研究的热点，而且随着异构的网络环境和组织环境的发展，分布式 workflow 系统越来越注重在实践中的应用，如基于分布式 workflow 的 CMS 应用集成、基于分布式 workflow 的企业集成建模方法、以及面向生产经营过程的资源仿真优化等。随着分布式 workflow 技术的进一步发展，它必将在提高企业竞争力方面做出更大的贡献。

参考文献：

- [1] ALONSO G, AGRAWALD, ABBADIEA, et al. Functionality and limitations of current workflow management systems 1997 [EB/OL]. <http://www.almaden.ibm.com/cs/exotica/wfnsys.ps>
- [2] TOMLISON C, ATTIE P, CANNATA P., MEREDITH G, SHETH A, SINGH, M, and WOELK D (1993). Workflow Support in Camot Bulletin of the Technical Committee on Data Engineering [J], 16 (2). IEEE Computer Society.
- [3] MCCARTHY, D, SARN S. Workflow and Transactions in InConcert Bulletin of the Technical Committee on Data Engineering [J]. 16 (2). IEEE Computer Society (1993).
- [4] MEDNAMORA R, WONG H, FLORES P. Action Workflow as the Enterprise Integration Technology. Bulletin of the Technical Committee on Data Engineering [J]. 16 (2). IEEE Computer Society (1993).
- [5] SHETH A. On Multi-system Applications and Transactional Workflows, Bellcore's projects PROMP and METEOR. Collection of papers and reports from Bellcore (1994).
- [6] 范玉顺. 工作流管理技术基础 [M]. 北京: 清华大学出版社, 2001.
- [7] 张宁, 文静华, 余霏. 基于过程模型的分布式 workflow 引擎的研究 [J]. 贵州大学学报 (自然科学版), 2007, (3): 290 - 293.
- [8] LISMA L and DHAGMONT. A performance evaluation of the mobile agent paradigm. In Proceedings of the Conference on Object - Oriented Programming, Systems, Languages, and Applications, pp. 306 - 313, 1999.
- [9] NICHOLAS R JENNINGS, MICHAEL J WOOLDRIDGE. Applications of Intelligent Agents [J], Agent technology: foundation, applications, and markets, Springer, Germany, 1988, pp. 3 - 28.
- [10] DWENGER, ARPROBST. Intelligent Agents in Financial Services [J]. Agent technology: foundation, applications, and markets, Springer, Germany, 1988, pp. 303 - 325.
- [11] 李红信, 范玉顺. 分布式 workflow 执行服务的设计与实现 [J]. 计算机工程与应用, 2003, (24): 118 - 120.
- [12] WEISSENFELS J, WODTKE D, WEIKUM G, DITTRICH A. The Mentor Architecture for Enterprise - wide Workflow Management University of Saarland, Department of Computer Science, 1997.
- [13] GALONSO, CMOHAN, R GINTHOR, D AGRAWAL, A El Abbadi, M KAMATH. Exotica/EMQM: A Persistent Message - Based Architecture for Distributed Workflow Management Proc. IFIP Working Conf on Info Sys for Decentralized Organizations, Trondheim, 8/95, pp15.
- [14] BARBARA D, MEHROTA, S, RUSNKEWICZ, M. NCAS: A Computation Model for Dynamic Workflows in Autonomous Distributed Environments [M]. Technical report, Matsushita Information Technology Laboratory (1994).
- [15] BARBARA D, MEHROTRA S, RUSNKEWICZ M. NCAS: Managing Dynamic Workflows in Distributed Environments [J]. Journal of Database Management, Vol 7, No 1, 1996.
- [16] JACQUES WANNER, EDMUNDO R MADEIRA. A Fully Distributed Architecture for Large Scale Workflow Enactment [J]; 2003. 3; pp14.