

网格环境下一个基于服务的工作流管理系统

李志清, 傅秀芬, 苏辉贵, 蒋明亮, 肖树腾

(广东工业大学 计算机学院, 广东 广州 510090)

摘要: 网格环境是一个分布、动态、异构的环境, 在这样的环境里, 对服务进行很好的管理是非常重要的. 本文基于网格服务, 提出了一个分布式的工作流管理系统, 研究了系统中的组件和一个进程的运行过程, 并介绍了该系统的实现, 为网格服务的商业化提供了有力的支持.

关键词: 网格计算; 网格服务; 工作流管理系统

中图分类号: TP393

文献标识码: A

文章编号: 1007-7162(2006)04-0088-05

引言

网格计算作为一个重要新领域, 与一般的分布计算有很大的区别, 它所强调的是大规模的资源分享, 创新性的应用, 并且在某些情况下, 获得高性能的服务^[1]. 网格计算已经成为一个主流技术, 能分享大规模分布的资源并使系统综合化. 由 Globus Alliance 提出的开放网格服务体系结构 (OGSA)^[2] 目的是为了处理和集成分布在异构, 动态和复杂网格环境下大规模的资源. 在 OGSA, 任何事情都看作是一个网格服务. 因此, 无论是企业的, 还是研究机构的研究人员, 都致力于把网格服务组成为大的应用和信息系系统^[3]. 工作流是一个业务流程的自动化, 在整体或部分上, 在信息或任务期间, 根据一套程序规则来制定从一个参加者到另一个参加者的行动^[4]. 由于工作流技术在进程管理中的广泛应用, 它成为了处理服务合作, 服务构成和服务交互的一个重要方法, 也成为了一个管理进程和资源的重要方法. 一个工作流管理系统 (WFMS) 是一个通过利用软件运行在一个或更多工作流引擎上定义, 创建和处理工作流的系统^[4]. 目前, 有很多问题要依靠网格工作流和服务流来解决. 因此, 提出了许多语言和标准, 例如 BPEL4WS, GSFL^[5] 等等. 但是, 它们中的大多数都强调在对关系的描述和在服务之间交互作用, 往往忽视了工作流管理系统的构造. 事实上, 在开放和动态网格环境里怎样构造一个强大的高效率的工作流管理系统也同样是一个很重要的问题. 一个工作流管理系统不仅是一个管理资源的工具, 而且也是网格环境中的一个服务群. 网格为共享和合作资源提供了机制, 一个工作流管理系统的功能就是要在大范围里共享和合作资源. 本文提出了在网格环境里构造一个基于服务的分布式工作流管理系统的一个方法. 在一个传统工作流管理系统里, 多数实用组件都是抽象的, 并且被提升为更高水平的网格提供服务. 因此, 构造一个基于服务分布式的工作流管理系统, 目标是要把所有的功能服务都组成为一个整体.

1 S-WFMS 的概述

由 WFC^[4] 提出的传统工作流参考模型里, 工作流管理系统包括了进程定义工具、管理和

收稿日期: 2005-10-18

作者简介: 李志清 (1981-), 男, 在读研究生, 主要研究方向为网络软件及多媒体技术, 网络技术, 网络安全.

监控工具、客户应用以及工作流引擎. 在同一个工作流管理系统里, 进程是由进程定义工具来定义, 并且用工作流引擎来执行. 任务由终端用户执行或应用, 而进程仅仅能被引擎来执行, 引擎是由具体用户和具体应用来获取.

在网格环境里运用工作流管理系统已不再受到它们具体的地点制约. 任务是由网格服务执行的. 在网格服务有了一套标准接口之后, 它们能根据标准由任何应用程序来获取. 因此, 一个进程的执行不会受到工作流引擎地点的限制, 进程能在任何地方设计, 任何地方存储和执行. 用户能选择分布在网格环境里的任何可利用的工作流引擎服务来开始执行一个进程. 在一个进程执行期间, 它可能在不同的工作流引擎服务之中进行转移来达到好的装载平衡, 更低的通信费用, 遇到缺点时更容易修复等等. 甚至, 一个进程能同时在几个工作流引擎服务中执行, 从而达到更高的效率.

根据工作流参考模型, 一个工作流管理系统的主要功能是可以划分成几个相互独立的客户组件和网格服务, 其中包括进程定义工具, 管理和监控工具, 进程定义存储服务, 工作流引擎服务和 workflow 事例存储服务. 进程定义工具和管理监视工具是配置在客户端的客户工具. 其它 3 个网格服务是由分布在网格环境里所有的组织和企业所提供的. 并且, 工作流 Metadata 存储服务和工作流网格注册和发现服务也是需要的. 如图 1 显示了在网格环境里分布式工作流管理系统的结构. 在这一结构中, 每一个组件的服务和工具的详细信息将在下一部分介绍.

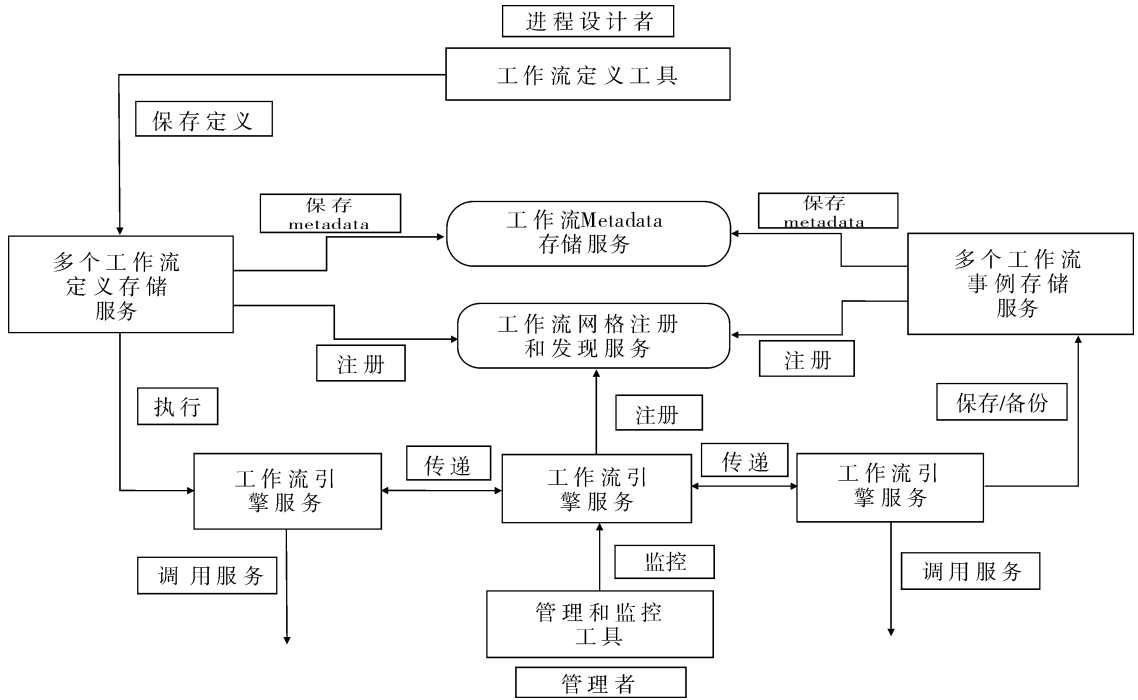


图 1 SWFMS 的结构

2 SWFMS 的组件

2.1 工作流定义工具

工作流设计者利用这个工具在客户端定义和设计进程. 设计的结果形成一个描述的文件, 详细请查看 GSFL^[5].

2.2 管理和监视工具

当进程被执行的时候,进程的管理者就会使用这个工具监控和管理所执行的过程.例如, `pause/stop/restart` 执行, 转移到另一个工作流引擎服务上执行.

2.3 工作流定义存储服务

这类网格服务被用来存储进程定义文件.它所支持的操作有:增加,修改,删除和获取进程定义的文件.在基于服务的分布式工作流管理系统中,一个进程定义文件可能比存放用不同的存储服务的备用空间还大.

2.4 工作流事例存储服务

它被用于存储进程事例信息.一个进程事例也可能比存放用不同的工作流事例存储服务的备用空间还要大.当在执行过程中发生错误时,备份被用来修复错误.这个服务支持的操作有:增加,更新和获取进程事例信息.

2.5 工作流引擎服务

在工作流管理系统中,这是核心组件网格服务.它是负责进程事例的执行,并且支持的操作有:解析进程定义文件,引起进程事例,调用外部的网格服务和应用程序,处理数据/信息流等.

2.6 工作流网格服务注册与发现服务

在网格环境里,这项服务与其它索引或目录服务是相似的.它被用来存储工作流网格服务的信息,包括服务提供者,服务 QOS 服务价格,服务存取方式等.这项服务也负责服务询问.它支持的操作有:登记,删除,更新和查询等.

2.7 工作流 Metadata 存储服务

这项服务用来存储工作流 metadata 例如,一个进程定义文件的位置,哪个工作流引擎服务执行进程,对应的工作流事例被存放的位置等.

3 在 S-W FMS 里设计 执行 / 管理进程

在 S-W FMS 里,任何操作,例如,设计 执行 / 管理一个进程都需要几个相关的网格服务组件和工具共同来完成.

3.1 设计一个进程

首先,用户在客户端用一个工作流定义工具设计一个进程.第二,设计者为了一个或者更多想要工作流定义存储服务,要求工作流网格服务注册和发现服务来存储进程定义文件.最后,工作流定义存储服务保存了从进程定义文件的位置到工作流 metadata 服务的 metadata 信息.

3.2 执行一个进程

首先,用户通过工作流 metadata 服务发现了目标进程定义文件的位置.第二,用户从目标工作流定义存储服务中检索进程定义文件.然后,通过在工作流网格服务注册和发现服务上查询来选择一个工作流引擎服务,用该工作流引擎服务来执行工作流.之后,工作流引擎服务把 metadata 保存到工作流 metadata 服务.工作流引擎服务再引起一个进程事例,并把它保存到一个已选择的工作流事例存储服务.最后,工作流引擎服务开始执行这个进程.在执行过程中,工作流引擎服务还会调用外部的网格服务或应用程序执行进程中的任务.

3.3 管理一个进程

首先,用户通过在工作流 metadata 服务上查询获得目标工作流引擎服务执行进程和工作流事例存储服务.第二,用户运用工作流引擎服务的管理和监控接口获取进程事例的相关数据或

改变进程事例中的一些属性. 其结果可以在客户端的管理和监控工具中显示出来.

4 S-WFMS的实现

一个基于服务的 WFMS是由独立工作流网格服务和分布在网格环境里的其他工作流组件组成的. 与建立一个传统的工作流管理系统相比, 它使用了服务构造的方法, 更容易建立一个基于网格服务的 WFMS. 建立一个基于服务的 WFMS的重点是对工作流网格服务的理解. 事实上, 工作流网格服务, 例如工作流引擎服务可以根据网格服务标准, 通过对现有传统工作流组件进行一定的修改和更进而实现. 也就是说, 可以增加一些标准接口到那些现有组件, 在组件中引入网格存储服务 and 注册发现服务. 下面对本文所提出的 S-WFMS结构中的重点部分工作流引擎服务进行详细的介绍.

工作流引擎服务应该支持以下接口:

- 1) 开始进程接口. 这个接口在开始执行一个进程时被调用. 输入参数的格式在 GSFL^[5]里的工作流定义文件中, 这个进程的原始数据和其它的一些配置引起新的进程实例.
- 2) 查询进程接口. 这个接口用来检索进程的状况, 相关数据和进程实例的其它属性. 参数就是这个进程实例的标识符和名字.
- 3) 转移进程接口. 当一个进程将被转移, 该接口用来发送进程实例的执行信息. 这个参数是有关一个进程实例和目的地引擎服务的信息. 这个接口也接受进程实例的信息.
- 4) 同步数据接口. 当一个进程由几个工作流引擎一起执行时, 它将维护进程数据的连贯性和一致性. 参数包括了数据标识符, 数据价值, 修改时间等等.
- 5) 调用外部服务接口. 当执行任务时, 它被用于调用外部网格服务或网格服务.

为了使一个工作流引擎服务运用在功能接口之上, 我们能利用传统工作流参考模型中对工作流引擎的接口定义. 下面我们对工作流引擎服务的主要接口, 在已有的工作流管理应用程序编程接口 (WAPI) 的基础上, 说明工作流引擎服务的主要接口还应该扩展的功能, 用来实现所构造的 S-WFMS.

入口进程定义. 在一个进程开始之前, 这个进程定义文件应该被放进工作流引擎中. WAPI提供的接口 1 具有这个功能. 然而, 这个接口只支持 XPDL/WPDL 文件, 我们应该扩展这个接口完成支持 GSFL 文件的功能. 更进一步地说, 这个接口应该被扩展到有能力访问工作流定义存储服务, 以便得到目标 GSFL 文件.

开始一个进程. WAPI 的接口 2 提供了这一操作. 因此, 它也能被用来在工作流引擎服务里开始一个进程. 但这个接口还应该扩展成具有访问工作流实例存储服务以便存放实例, 以及访问工作流 metadata 存储库服务和工作流网格注册和发现服务.

调用外部网格服务. WAPI 的接口 3 只支持调用外部网络的应用. 事实上, 每一个网格服务都可以认为是一种特殊的应用. 因此, 我们能增加一些服务的调用代码到这个接口中, 使得这个接口支持调用网格服务.

转移进程. 在 WAPI 的接口 4 中, 共用接口定义提供了这种功能.

执行进程的一致性和维护数据的连贯性. 这一特性不是所有 WFM 共用模型都支持. 但有些模型和产品提供了这种功能.

从上述比较来看, 一个工作流引擎服务所要求的特性可以在传统工作流参考模型上, 通过扩展参考模型的接口来实现. 这样, 我们就可以根据网格服务标准, 把一个带有一定扩展的传统工作流引擎运用到一个网格服务中, 实现一个工作流引擎服务. 我们对 S-WFMS 的其他组件, 比

如, workflow 定义工具, 管理和监视工具, workflow 定义存储服务等, 也利用传统的工作流参考模型中已有的组件基础上, 进行进一步功能的扩充和完善. 这样, 我们对 S-WFMS 的构造和开发, 既大大减少了开发的时间和成本, 在质量、安全性等各方面也有一定的保证.

5 总结和展望

随着网格技术的发展, 它受到了工业领域和研究领域越来越多的关注^[6]. 网格环境是一个分布、动态、异构的环境^[7], 在这样的环境里, 对服务进行很好的管理是非常重要的. 许多研究者都提出运用 workflow 技术来处理在网格环境下对服务的管理问题. 文中我们提出了基于 workflow 管理系统的一个构架. 该构架有很多好的特征, 例如高效率, 更好的容错能力, 更好的装载平衡性和可靠性等. 与传统的工作流管理系统比较起来, 一个基于服务的分布式 workflow 管理系统能支持大范围的应用, 更好的装载平衡能力和容错能力, 更高的执行效率. 与其他的分布式 workflow 管理系统比较, 它适应大范围的应用和支持分布在 Internet 上的网格服务, 并且它能更方便处理在工作流上的一些变化, 更能适应动态环境. 我们认为, 随着网格技术发展得越来越普遍, 构造一个基于服务的分布式 workflow 管理系统, 在企业处理它们的进程和工作流管理进行深入的研究上, 都有着重要的作用和意义.

参考文献

- [1] Ian Foster, Carl Kesselman, Steven Tuecke. The anatomy of the grid: Enabling scalable virtual organizations[J]. International Journal Supercomputer Applications, 2001, 15(3): 200-222.
- [2] Ian Foster, Carl Kesselman, Jeffrey Nick, Steven Tuecke. The physiology of the grid: an open grid services architecture for distributed systems integration[C/OL]. In Proceedings of the 5th Global Grid Forum Workshop (GGF5), Edinburgh, Scotland (2002-07). [2005-09-12]. <http://www.globus.org/>.
- [3] K. Sivashanmugan, J. A. Miller, A. P. Sheth. Framework for semantic web process composition[J]. International Journal of Electronic Commerce, 2004, 12(3): 212-224.
- [4] D. Hollingsworth. Workflow management coalition. The workflow reference model[M/OL]. Document Number WFM-C-TC00-1003, Brussels, 1994 [2005-09-12]. <http://www.wfmc.org/standards/docs/tc003V11.pdf>.
- [5] Sriam Krishnan, Patrick Wagstrom, Gregor von Laszewski. GSFL: A workflow framework for grid services, 2002 [2005-09-12]. <http://www.unik-gbbs.org/cog/papers/gsf-paper.pdf>.
- [6] Ian Foster, Carl Kesselman. 网格计算[M]. 金海, 袁平鹏, 石柯译. 北京: 电子工业出版社, 2004: 2-18.
- [7] 都志辉, 陈渝, 刘鹏编写, 李三立院士审阅, 网格计算[M]. 北京: 清华大学出版社, 2002: 1-12.

A Service-based Workflow Management System in Grid Environment

LI Zhi-qing FU Xiu-fen SU Hu-gui JIANG Ming-liang XIAO Shu-teng

(Faculty of Computer, Guangdong University of Technology, Guangzhou 510090, China)

Abstract In grid environment, it is important to manage services in a distributed, dynamic and heterogeneous environment. This paper proposes a distributed workflow management system (WFMS) based on grid services, investigates components and the operation of a process in S-WFMS, and introduces an implementation of S-WFMS. Thus, the system provides powerful support for commercial grid application.

Key words grid computing; grid service; WFMS