

面向数字化档案管理系统 SaaS 应用研究

吴振宇¹, 吴 健², 张 涛¹

(1 西北工业大学 软件与微电子学院, 陕西 西安 710072; 2 西北工业大学 计算机学院, 陕西 西安 710072)

摘 要: 软件即服务(Software-as-a-Service, SaaS)已经变成十分流行的软件模型, 租户可以通过互联网访问 SaaS 化的服务, 并且通过 SaaS 服务提供商租用所需服务. 以某省数字化档案馆的建设为背景, 结合了 SaaS 服务模型支持多租户, 灵活可定制的特性, 提出一种基于 SaaS 的数字化档案管理系统架构, 该架构不仅支持多租户访问, 而且支持服务组件的定制, 有利于解决日益复杂的档案管理问题, 也有利于避免各地市档案馆建设的重复投资. 重点对支持多租户的存储模型的构建, 以及档案业务逻辑在线定制模型进行了研究, 提出了一种在线的动态代理模型—ActiveProxy 模型, 应用效果良好.

关键词: SaaS; 可定制服务; 多租户存储; 数字化档案馆

中图分类号: TN402

文献标识码: A

文章编号: 1000-7180(2014)04-0013-05

The Application and Research for Digital Archives Management System Based on SaaS

WU Zhen-yu¹, WU Jian², ZHANG Tao¹

(1 College of Software Engineering, Northwestern Polytechnical University, Xi'an 710072, China;

2 College of Computer, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: SaaS (Software as a Service) has been becoming the very popular software mode; the tenants can get the SaaS service from the Internet, and rent the needed service through the SaaS service provider. This paper proposes the Digital Archives Management Architecture based on the SaaS through combining the supporting multi-tenant, service customization and the requirement of the Digital Archives. The Architecture has the peculiarity that supports the visit from various tenants and is easy to be customized, in favor of giving solution to the archives management logic, and prohibiting redundant costs in various cities. This paper also study the storage model supporting the multi-tenants, and the archives customization model to promote the construction of digital archives.

Key words: SaaS; customization service; multi-tenant storage; digital archives

1 引言

SaaS 作为一种新的软件服务模式, 越来越多的应用到企业信息化平台的建设当中, 目前 SaaS 应用中多租户的数据隔离方式以及业务逻辑的定制模型日益成为国内外研究的热点. 文献[1]中提出了 SaaS 模式下多用户数据模型, 文献[2]中进一步讨论了目前主要的集中多租户数据管理方式, 包括元数据, XML 等. 文献[3]中基于数字化图书馆, 使用

XML 技术对传统的文本格式进行了转化, 存储. 但是都是针对 SaaS 模式下数据的存储方式进行探讨, 数据隔离方式的讨论不够深入. 文献[4]中提出了用户模板的概念, 将元数据根据用户模板进行划分, 以便提供个性化的服务. 但是如何进行业务逻辑定制, 如何维护组件间的业务流程, 并没有经过验证. 经过对以上文献论证和分析后认为, 多租户存储模型的数据隔离和业务逻辑的定制模型还有很多的问题需要研究, 本文主要对数据隔离的关键技术和在线可

收稿日期: 2013-06-30; **修回日期:** 2013-08-19

基金项目: 国家自然科学基金-青年科学基金课题(61103003); 陕西省自然科学基金(2011JQ8008); 西北工业大学研究生种子基金(Z2012189)

定制模型进行了探讨和研究。

2 基于多租户可定制模型的数字档案馆架构设计

面向多租户存储和可定制服务是 SaaS 模式下数字化档案馆构建的关键因素,其目的是从逻辑上耦合各类档案资源并通过一致性的访问控制策略向多个租户提供统一的访问接口。

本文提出了面向多租户的可定制的档案管理框架,如图 1 所示,档案管理框架划分为基础设备层、

多用户平台层,最底层的基础设备层,负责提供各种数据库服务器,主要用于维护业务数据库、租户数据库等;在此之上为多用户平台层,负责多用户运行环境的部署和维护,这一层分为两个子层:上层也负责维护组件容器和线程池,主要是通过线程池为每个租户分配不同的线程接受租户的请求,与预先定义的服务组件进行交互。下层就是资源管理层,资源管理的基本功能是接受来自租户的请求,并将特定资源向资源服务进行注册,与此同时,这一层负责资源的调度,使请求资源的任务得以即时地执行。

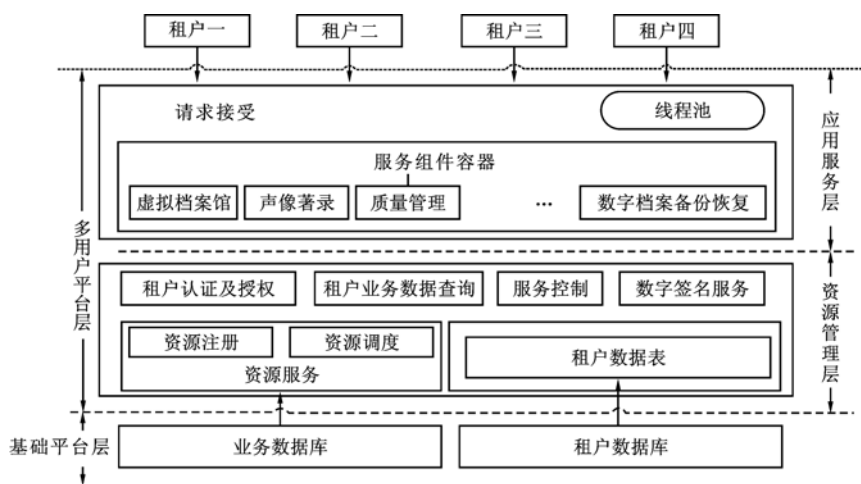


图 1 基于多租户可定制模型的档案馆数字化架构

3 面向 SaaS 应用的多租户存储模型构建

数据的隔离与共享策略是 SaaS 服务模型设计的关键因素,在大多数 SaaS 化应用中,许多企业都使用同一个应用实例;应用实例必须只允许每个租户自己的成员访问该租户的数据,能够让多个租户共存于相同的应用实例中,同时不会破坏数据的独立性,具备这种能力的应用程序就可以称之为多租户应用程序。这种应用实例的关键技术在于多租户存储模型的设计。目前实现多租户存储模型主要有以下三种方案^[5-6]。

(1)对每个租户使用不同的数据库。

(2)共享数据库,分离数据模式。

(3)共享数据库和数据模式。在具体的档案数据表中对不同的租户进行隔离。

通过对三种方案的分析,可以发现,由于档案馆数据规范、标准化程度高、数据结构一致性好、数据独立性低,采用共享数据库与数据模式的方案较为适宜。

在这种方案中,为了实现租户的数据隔离,首先需在相关数据模型中增加租户的角色,以标志该数

据模型所属的具体租户;然后在应用系统的功能设计中,向每个表和视图添加一个租户标识符;最后,建立租户权限配置访问控制列表,并在相关操作(如查询、修改、删除和统计)执行前,根据租户标识符,判断租户的权限,以保证租户间的数据隔离和数据安全^[7]。

下面以数字化档案馆中的虚拟档案室为例进行说明。虚拟档案室是针对各种档案权属机构建立的数字化档案平台,档案权属机构可以根据自身需求通过数字化档案馆建立自己的档案管理平台。

图 2 所示即为虚拟档案室的数据模型,虚拟档案室是档案的逻辑存储机构,由其档案权属机构(区县档案馆,中小企业)进行管理。

此模型可以被定义为一个 4 元组。VirArchiveRoom = (User, Dept, VirRoom, Archives);

VirRoom = {RoomID, RoomName, DeptID} 是虚拟档案室的信息,RoomID 是虚拟档案室的唯一标识,DeptID 是权属机构的标识。

Archives = {DWCode, RoomID} 是档案的数据信息,DWCode 是档案定位符,是整个数字化档案馆中档案数据的唯一标识,RoomID 是指所属虚

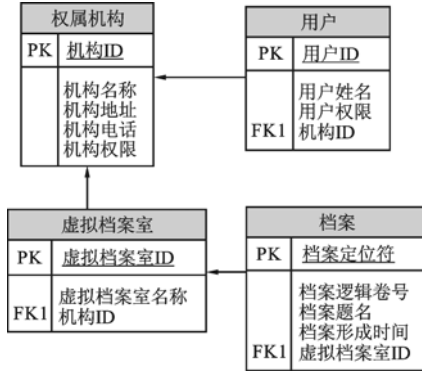


图2 虚拟档案室的数据库数据模型图

拟档案室的标识。

在本例中,档案数据对象可以被定义为一个二元组 $DATA = (VirRoom, Archives)$, $DATA$ 是要进行隔离的数据对象。

$Dept = \{DeptID, DeptName, DeptAuthority\}$ 是权属机构基本信息,其中 $DeptID$ 唯一标识用户, $DeptAuthority$ 为权属机构的基本权限,只是针对本机构的档案一些访问,查找操作的权限

$User = \{UserID, UserName, Authority, DeptID\}$ 是用户的基本信息,其中 $UserID$ 是唯一标识用户, $UserName$ 为用户名, $Authority$ 为用户权限,用户权限在其权属机构之上可以拥有审批,删除档案的权限,通过申请甚至可以拥有访问其他机构的权限。

在此例中, SaaS 模式下的租户的概念并不仅仅指用户,在这里通常包括托管档案的机构(如各地市的档案馆,中小企业等等)及其所属用户。租户通常被描述为如下键值对。

$$TenantID = \langle UserID, DeptID \rangle$$

其中 $DeptID$ 下可以对应多个 $UserID$, 即 $Dept$ 和 $User$ 之间存在一对多的对应关系。用户首先登陆虚拟档案室,系统通过当前会话获取当前用户的 ID 进而该用户所属的档案机构 ID ($DeptID$)。

这样,档案数据就在用户和机构两个层级上进行了隔离。不同的租户访问时,权属机构和档案对象数据生成一张视图,其关系代数如下。

$$\Pi_{DeptID}(Dept) \bowtie \langle (\Pi_{RoomID, DeptID}(VirRoom) \bowtie \langle \Pi_{DWCCode, RoomID}(Archives) \rangle) \rangle$$

这样就实现了在用户和机构两个层级上实现了对档案对象数据的隔离。

4 业务逻辑可定制模型构建

在 SaaS 模式下,数字化不但要支持租户隔离,而且富于扩展性。多租户要根据环境的不同,比如不

同的工作流程、不同的服务层级等进行功能扩展。结合国内外研究现状,本文提出了一套业务逻辑在线可定制的动态代理模型——ActiveProxy 模型。该模型主要包括以下几个核心部分。

(1) 定制组件管理。即服务提供商对可定制服务列表中组件进行管理,根据业务调研,服务提供商可以对定制组件进行封装和压缩。其定义如下:

$S(\text{服务组件}) = \{C_1, C_2, C_3, C_4 \dots\}$, S 中包含了所有根据业务调研开发的服务组件。

(2) 定制规则管理。即服务提供商对可定制组件列表中的组件之间的依赖关系进行管理等价于业务规则的预定义。其定义如下:

$$R(\text{服务组件}) = \{(C_1 \rightarrow C_2), (C_1 \rightarrow C_3), (C_1 \rightarrow C_4) \dots\}.$$

R 中包含了所有由服务提供商预先定义的业务规则,以便于验证租户申请的业务流程符合该规则。例如 $C_1 \rightarrow C_2$ 中 C_2 是 C_1 的直接前趋, C_1 发生之前 C_2 已经发生。

(3) 服务规则的管理及其验证。服务验证的作用是根据服务规则对租户提交的请求进行验证。租户提交的组件配置关系如下:

$P_S = \{C_1, C_2, C_3 \dots\}$, P_S 为该租户请求组件的集合。

$P_R = \{(C_1 \rightarrow C_2), (C_1 \rightarrow C_2 \rightarrow C_3) \dots\}$, P_R 描述了该租户的具体业务流程,表示为一个包含了请求组件依赖关系的集合。

业务流程分顺序和并发两种,分别表述如下。

图3中组件之间的依赖关系分别表示 $P_R(a) = \{(C_i \rightarrow C_j)\}$, $P_R(b) = \{(C_k \rightarrow C_j \rightarrow C_i)\}$ 。图3(a)中 C_i 是 C_j 的唯一前趋, C_j 发生前 C_i 已经发生。图3(b)与图3(a)中的依赖关系类似,描述了三个组件间的关系。

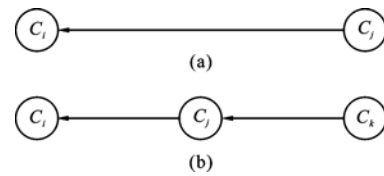


图3 顺序业务流程的依赖关系

如图4所示,有两种位置关系,有一组并发服务组件有相同的直接后继(图4(a)所示),或者有一组并发的服务组件有相同的直接前趋(图4(b)所示)。并发工作流程的依赖关系通常以 n 元组的形式出现。

$$P_R(a) = \{(C_k \rightarrow C_i, C_k \rightarrow C_j)\},$$

$$P_R(b) = \{(C_j \rightarrow C_i, C_k \rightarrow C_i)\}.$$

组件(即 P_S) 的验证算法如下:

IF $P_S.size \leq S.size$

THEN FOR($j=1$ to $P_S.size$) DO

IF $P_S.C_j \notin S$

RETURN FALSE

ENDIF

ENDFOR

RETURN TRUE

ENDIF

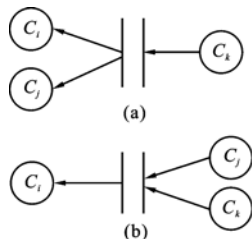


图4 并发业务流程的依赖关系

针对租户提交的组件间依赖关系的验证较为复杂,由于业务规则中的流程描述均为二元组,首先需要对租户提交的依赖关系进行映射和转化.如表1所示.

表1 组件依赖关系和结构化集合之间的映射

组件间的依赖关系	对应的集合
$PR1 = \{(C_k \rightarrow C_j \rightarrow C_i \rightarrow \dots \rightarrow C_n)\}$	$Result1 = \{C_k \rightarrow C_j, C_j \rightarrow C_i, \dots, C_{n-1} \rightarrow C_n\}$
$PR2 = \{(C_k \rightarrow C_i, C_k \rightarrow C_j)\}$	$Result2 = \{C_k \rightarrow C_i, C_k \rightarrow C_j\}$
$PR3 = \{(C_j \rightarrow C_i, C_k \rightarrow C_i)\}$	$Result3 = \{C_j \rightarrow C_i, C_k \rightarrow C_i\}$

根据映射转换得到的集合 $Result_1, Result_2, \dots, Result_N$ 以及预先定义的业务规则 R ,如果存在如下逻辑关系,则业务流程验证通过,否则业务流程验证失效.

$$\forall r(r \in \bigcup_{i=1}^n Result_i \rightarrow r \in R);$$

根据此模型,结合档案馆的具体建设需求,业务流程定制过程可以分为以下五步,如图5所示.

(1) 服务组件的初始化和业务规则的预定义.档案馆的超级管理员根据预先定义的服务内容进行分类,将涉及不同内容的服务组件进行封装.同时可以通过定义组件之间的依赖关系定制不同的业务流程.服务组件集合 S 以及组件之间依赖关系集合 R 如下.

$S(\text{服务组件}) = \{C_1(\text{档案著录}), C_2(\text{档案鉴定}), C_3(\text{档案审批}), C_4(\text{档案销毁}) \dots\}$

$R(\text{服务组件}) = \{C_2(\text{档案鉴定}) \rightarrow C_1(\text{档案著录}), C_3(\text{档案审批}) \rightarrow C_1(\text{档案著录}), C_3(\text{档案审批}) \rightarrow C_1(\text{档案著录}), C_4(\text{档案销毁}) \rightarrow C_3(\text{档案审批})\}.$

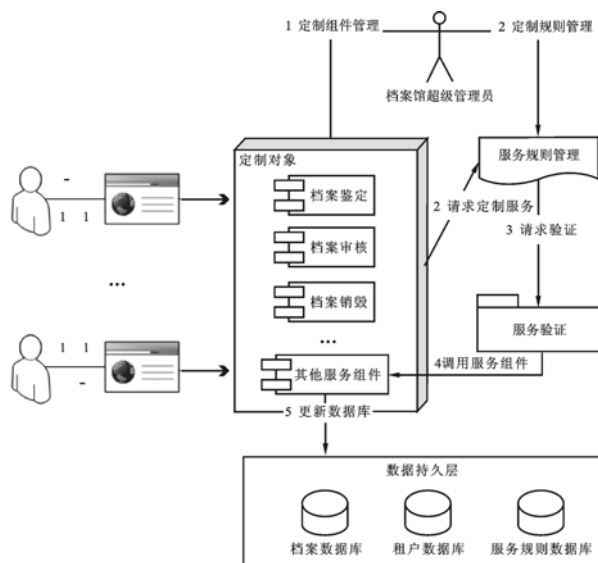


图5 ActiveProxy 模型架构图

审批)}.

(2) 请求定制服务. 区县档案馆等机构可以通过网页或者其他可视化的终端,根据自己的需求,对不同的组件进行选择和组合来构造自己的档案管理系统,同时还可以根据自身业务规则对组件之间的依赖关系进行配置.例如某区县档案馆提交如下请求.

$P_S = \{C_1(\text{档案著录}), C_2(\text{档案鉴定}), C_3(\text{档案审批})\}$, P_S 为区县档案馆请求组件的集合.

$P_R = \{C_2(\text{档案鉴定}) \rightarrow C_1(\text{档案著录}), C_4(\text{档案销毁}) \rightarrow C_3(\text{档案审批}) \rightarrow C_1(\text{档案著录})\}$, P_R 描述了该档案馆的具体业务流程.

(3) 请求验证. 服务对象模块会收集数据库中可定制的服务信息以及其他信息,并将这些信息提交给验证引擎.服务验证引擎接受请求验证信息,根据服务规则中的配置信息进行分析和比对.根据上文给出的比对算法, P_S 通过请求, P_R 转化为 $Result_1 = \{C_2(\text{档案鉴定}) \rightarrow C_1(\text{档案著录})\}$, $Result_2 = \{C_4(\text{档案销毁}) \rightarrow C_3(\text{档案审批}), C_3(\text{档案审批}) \rightarrow C_1(\text{档案著录})\}$. 根据 $Result_1, Result_2$ 和 R (服务组件) 三个集合, P_R 业务验证流程通过.

(4) 调用服务组件. 当该区县档案馆完成整个定制流程,服务验证引擎便开始验证业务逻辑的正确性.如果验证成功,验证引擎向对象模型中的相关组件发出指令,完成业务逻辑.

(5) 更新数据库. 验证通过以后,档案馆超级管理员可以根据该区县档案馆的需求修改相应信息,但是为了不使其他租户受到影响,定制对象模块会将定制信息保存在相应的数据库中.

5 结束语

本文以省级档案馆的数字化建设为背景,设计了面向多租户可定制的管理系统模型,并从多租户存储模型和业务逻辑的可定制框架两个方面对该模型进行分析和阐述.该模型以较为低廉的成本,支持了各级档案馆的租户,同时可以验证业务逻辑的正确性,通过组件化的服务防止异常情况的发生.

进一步的研究工作主要集中在如何扩展服务规则管理模块,包括如何定义提供数据服务的方法和制定标准 API 接口,以便支持一定范围内的租户自定义处理逻辑.

参考文献:

- [1] Zhang Xiangyu, Ai Jing, Wang Zhongyuan, et al. An efficient multi-dimensional index for cloud data management [C]//Proceedings of the CloudDB. HongKong, China, 2009:17-24.
- [2] Shi Yuliang, Luan Shuai, Li Qingzhong, et al. A multi-tenant oriented business process customization system [C]//New Trends in Information and Service Science.

(上接第12页)

残留信息的相对重要程度 α 、预见值的相对重要程度 β 和信息素的挥发系数 ρ ,体现了相关信息痕迹、预见度及挥发对蚂蚁决策的相对影响.为了进一步分析各参数对解的影响,我们选用不同的 α 、 β 、 ρ 取值.实验表明, α 、 β 、 ρ 的取值变化均会影响问题的求解结果,多次运行后得出本文的最佳参数设定为 $\alpha=1$, $\beta=2$, $\rho=0.5$,此时得到的实验结果最优,且收敛速度最快,具有较强的全局寻优能力.

实验结果表明本文针对云计算环境下的虚拟机负载均衡问题建立的模型是可解的,算法是有效的.由表1可以看到由改进的蚁群算法得到的结果更优,而且具有更快的收敛速度,算法的性能更优.

5 结束语

本文对云计算环境下的虚拟机负载均衡问题进行了研究,建立了数学模型,并针对其中特定约束条件进行了研究,最后通过蚁群算法进行了求解,对指导现实生活中的实际情况具有一定的理论意义.

参考文献:

- [1] Li Yajun, Yang Yuhang, Ma Maode, et al. A hybrid

Gyeonqiu, 2009.

- [3] 刘莹. 基于 XML 的数字图书馆应用模型设计[D]. 济南:山东大学, 2011.
- [4] 管永明. 多租户软件中动态个性化配置与定制技术研究[D]. 济南:山东大学, 2011.
- [5] 张坤,李庆忠,史玉良. 面向 SaaS 应用的数据组合隐私保护机制研究[J]. 计算机学报, 2010(11): 2044-2054.
- [6] 李晓娜,李庆忠,孔兰菊,等. 基于共享模式的 SaaS 多租户数据划分机制研究[J]. 通信学报, 2012, (S1): 110-120.
- [7] 孔兰菊,李庆忠,史玉良. 面向 SaaS 应用基于键值对模式的多租户索引研究[J]. 计算机学报, 2011(11): 2039-2047.

作者简介:

吴振宇 男,(1988—),硕士研究生.研究方向为电子服务、云计算.

吴健 男,(1962—),教授.研究方向为软件工程、服务计算、电子服务、分布式计算、语义 WEB.

张涛 男,(1976—),副教授.研究方向为软件安全性、航空网络软件技术、软件测试.

load balancing strategy of sequential tasks for computational grids[J]. Future Generation Computer Systems, 2009, 25(8): 819-828.

- [2] 王少峰,周忠,吴威. 一种面向分布式虚拟环境的分层迭代负载均衡算法[J]. 软件学报, 2008, 19(9): 2471-2482.
- [3] 汪永琳. 超节点结构 P2P 中负载均衡的信息索引机制[J]. 计算机工程与科学, 2009, 31(8): 107-109.
- [4] Amaro de Sousa, Dorabella Santos, Pedro Matos, et al. Load balancing optimization of capacitated networks with path protection[J]. Electronic Notes in Discrete Mathematics, 2010, 36(1): 1249-1256.
- [5] Dorigo M. Optimization, learning and natural algorithms [D]. Italy: Politecnico di Milano, Department of Electronics, 1992.
- [6] Dorigo M, Caro G Di, Gambardella L M. Ant algorithms for discrete optimization[J]. An artificial Life, 1999, 5(2): 137-172.
- [7] 邢文训,谢金星. 现代优化计算方法[M]. 2版. 北京:清华大学出版社, 2005.

作者简介:

马华伟 男,(1977—),博士,副教授.研究方向为车辆调度问题、运筹与优化.