

# 分布式计算机软件数据整合平台设计

张识达

(哈尔滨医科大学附属第一医院 信息中心, 哈尔滨 150001)

**摘要:** 为实现当前经济市场中各应用软件的整合, 提高我国各行业间个性化软件的开发能力, 加强企业间的科学化系统软件设立, 保证企业的高效运行, 以及质量和系统维护能力; 根据较为常用的分布式计算机系统, 提出一种软件数据整合平台的设计; 以商用软件整合平台设计为例, 以开发模式、服务、技术结构、功能为研究方向, 在企业正常运行时, 可以实现业务系统功能开发以及门户单点登录, 提高企业的个性化系统应用和经济效益; 本文根据以上基本需求提出该软件平台进行设计, 并在实际应用中分析出对于当前企业发展所提供的助力; 结果显示, 在当前经济发展状况下, 有效地使用软件数据整合平台可以很大程度上提高企业的科学化管理和有效的资源分配, 使企业在市场竞争中提高经济效益和社会效益。

**关键词:** 分布式计算机系统; 软件数据整合; 管理; 资源分配

## Design of Data Integration Platform for Distributed Computer Software

Zhang Shida

(First Affiliated Hospital of Harbin Medical University, Information Center, Harbin 150001, China)

**Abstract:** For the current economic integration market integration of the use of application software, improve the Chinese personalized software development, strengthen the enterprise's scientific system establishment, to ensure efficient operation of enterprises and the quality and ability of system maintenance. According to the commonly used distributed computer system, a software data integration platform is proposed. With the commercial software integration platform design as an example, the development mode, service, technology structure and function for the open direction, in the normal operation of the enterprise can develop the business portal system function and single sign on, improve the personalized system application and the economic benefits of the enterprise. In this paper, according to the above basic needs of the proposed functional software platform design, and in practical applications analysis of the current enterprise development provided by the help. The results show that in the current economic development situation, the effective use of data integration software platform can greatly improve the enterprise's scientific management and effective allocation of resources for enterprises to improve economic and social benefits in the market competition.

**Keywords:** distributed computer system; software data integration; management; resource allocation

## 0 引言

在当前计算机网络技术急速发展的情况下, 各类型应用软件的开发速度有着明显的加快, 并且加之市场经济的高速发展, 当前各行业和企业对于同和市场资源以及信息化资源的要求越来越大。在大数据时代下, 科学化管理以及有效的资源整合是当前企业稳定发展和持续进步的重要支撑。现在企业在运行时依靠有效的应用软件, 对人员管理、资源管理、数据管理、信息管理等多方面都有更高的要求。而随着数据时代的不断加深, 在经济市场中所流动的市场信息和数据流量都非常大, 各个行业企业间对于信息化的认识以及依赖强度也逐渐加大, 对于开发和应用维护各种用途的应用型软件的需求也逐渐增高, 这也决定了当前软件的开发管理, 以及后期运作时的维护工作需要更高的要求。在当前的应用软件使用中, 大多数企业和行业都没有一个较为系统的软件数据整合办法, 各种用途的应用软件又因为特殊的应用需求所开发的框架不尽相同, 软件的管理出现很大的困难。并且各个软件系统的个性化架构、数据储存都相对于分散, 很难形成一个较为统一的管理模式<sup>[1]</sup>。

为此, 本文依据上述要求, 提出一种用于商用企业应用软件的数据整合平台, 利用分布式计算机系统的结构模式和运算能力, 设计出的软件整合平台, 为当前各企业间对于其应用的个性化软件以及功能性软件提供一种统一的管理办法。并且提高企业对于软件的科学化使用和有效的管理应用, 保证在大数据时代企业可以有有效的获取市场资源, 不断地提高经济效益以求稳定高效发展<sup>[2]</sup>。

## 1 分布式计算机系统

在进行这种软件整合平台的设计中, 采用分布式计算机系统是当今最有效的设计方向。其依照本身的无主从区分、信息共享能力、相互协作等特点, 对于各类应用软件的整合平台开发有很大的帮助。

### 1.1 分布式计算机系统的概要

分布式计算机系统是一种计算机配置方式, 其所面对的是计算机硬件配置以及功能型配置。又是一种多处理器的计算机系统, 在整个体系当中其构成依靠多个处理器通过互连网络而组成。分布式计算的结构是把原有系统内部的中央处理器任务通过分类组合, 将这些统筹后的任务条项分散给符合任务要求的处理器当中, 并且根据各个处理器之间的相互协作的办法, 共享系统的硬件设施以及内部软件。这种系统模式的采用可以大大的加快系统的整体处理能力和速度, 并且简化了主机的逻辑结构框架的模式, 对于当前的工业生产、自动化控制管理、

收稿日期: 2016-12-26; 修回日期: 2017-01-19。

作者简介: 张识达(1984-), 男, 黑龙江哈尔滨人, 助理工程师, 主要从事计算机软件设计方向的研究。

企业事业单位管理中都有着很高的应用价值。并且分布式计算机系统的成本较低,在维护工作上的操作也相对简单<sup>[3]</sup>,是当前计算机应用领域一个较为先进的发展方向。

### 1.2 分布式计算机系统的结构以及功能

在分布式计算机系统的结构向上包括了3个构成部分,通信结构、网络操作系统、分布式操作系统。在通信结构上主要作用是支持所有的计算机网络连接,并且可以提供所特定的分布式应用软件。在整个分布式的系统当中,每一个计算机都拥有自己的操作系统,这里的计算机就指的是各企业中的独立管理系统或是软件。而且这些系统和软件的种类都因其特定的工作需求而不相同,但是这些系统都应该支持同一个通信结构,这样就可以保证各个不相同的操作软件和系统可以通过网络进行连接。

网络操作系统主要的功能是提供相应的网络服务的,而这在本文中就是所需要设计与研究的软件数据整合平台。在分布式系统当中硬件环境是计算机网络,而系统当中的个人计算机则可以被看作是单个用户的工作站或者是服务器,而为保证多个工作站和服务器可以正常的联网运行以及有着统一的管理,则需要网络操作系统为其提供一个网络服务的支持<sup>[4]</sup>。

这些分散在各个企业或是各人计算机当中的软件都存在着一个公共的分布式操作系统。在整个分布式系统当中,每一台计算机都会共享一个公共的分布式操作系统。其组成的本分则是依靠一个内核为中心,加之多个系统功能模块和进程模块。位于分布式操作系统当中的计算机必须保存操作系统的内核数据,并且通过网络连接才可以做到对计算机系统当中的软件实现基本控制。

此外分布式操作系统还包括了分布式进程通信、分布式文件系统、分布式进程迁移、分布式进程同步以及分布式进程死锁等功能。分布式系统由于没有共享内存的办法,为此,系统的通信原则则需要按照通信协议的规定来实现。分布式文件系统则是根据了互联网连接的方式使得在每一个计算机上都可以得到共享文件的系统。并且还可以做到管理内核和文件之间的通信联系。分布式迁移则是代表着原机向目标机迁移数据和共享文件的功能,并且在目标机上依然可以正常进行系统操作。在分布式系统当中各个处理器不存在共享内存,为此在分布式系统进程当中则需要按照处理顺序的限号来进行分布式算法的排序。

## 2 在分布式系统下软件数据整合平台设计

由于当今市场中各类应用软件和管理软件种类繁多,并且软件的使用情况和面对的特定要求都不相同,在不同的行业和企业中都存在着自己的要求。为此,所需要的软件开发以及软件数据整合平台的设计开发都不尽相同。本文则根据目前大多数商用软件设计一种数据整合平台。这种整合平台可以保证企业内部信息化发展,并且根据不同的管理问题提供新的解决方案。而且利用这种平台企业可以根据自身的应用情况和实际的操作情况选择符合自身所需要的软件开发,强调的是企业的个性化软件应用和发展。在依靠这种软件数据整合平台的创新特性,企业可以快速的根据自身的发展现状完成所需软件的开发和部署工作。并且所开发处的软件简单易学,对于新软件的应用可以在很快的时间内便掌握软件的操作方式,而且依靠平台所提供的网络服务以及技术支持,对于后期使用中的维

护等工作都有着很高的便利性。

### 2.1 平台的模式设计

这个平台最大的应用特点在于应用数据的处理技术的革新。其中,元数据的处理和元数据库的开发是平台最为主要的部分。在平台当中对于所需要整合和管理的数据资源,其数据结构或是关联关系等信息资源都会储存在元数据库当中,当有应用需求的同时,数据库的处理器接收到访问请求对元数据资源进行解析工作,并且根据动态的数据访问指令形成上下统一的处理办法。而且这些元数据作为基础,根据相应的要求应用服务生成器则可以生成符合相应要求的软件操作页面或是结构。

#### 2.1.1 在线开发

在这个平台当中,软件的在线开发使其最大的特点。这种开发模式与以往的开放方式有所不同。开发人员只需要通过相应的账号或是指令登录到相应的操作网页当中进行软件的设计与开发工作。在整个开发的过程中可以及时观察在所开发阶段中此软件的应用方式和运行效果,对于部署环节有所简化,并且其方便和操作流程快捷。

#### 2.1.2 元数据程序开发模式

在这个平台设计的软件开发的过程中,其所采用的元数据开发模式有别于以往的代码编写模式。采用这种开发模式可以大大的减少软件开发过程中的难度,只需要对于元数据依靠应用要求进行条件编写,描述系统行为等方式。系统便可以根据所提供的条件进行模型构造以生成完整的操作软件或系统。

#### 2.1.3 自动生成调用接口

对于所开发出来的软件,所需要的调用接口的设计和开发在这种平台中无需开发人员特意设计。平台上的程序以及数据会自动根据 Web Service 的开放程序进行调用接口的生成。

### 2.2 平台所提供的服务

为满足当前商用的需求以及信息化自动化的发展,设计平台中所支持的服务模式包括3个主要的部分。信息服务功能、业务服务功能、界面服务功能。

在整个平台当中最为主要的就是信息服务的支持。在当前的数据时代中,信息资源是另一种有效资源,对于信息数据的掌握程度成为所开发软件的完整性和可实用性的标准。因此,所设计的分布式软件数据整合平台所得信息来源则是根据网络数据所提供的信息。并且将这些信息通过处理分类整合后储存在平台开发的数据库当中。为此,强调信息服务支持最为主要的就是平台的数据库开发工作。在这个平台的设计当中对于数据库的使用无需专业的数据库管理语言,平台数据库会根据所整合的资源模式进行数据开放导向,并建立起其整体结构模式。而且平台自身的数据库具备很高的隐私性以及访问权限的设定。企业如若需要某一特定的软件应用,其所购买的权限只可以在数据库当中访问特定的软件信息和资源,防止了软件的外泄以及市场信息的规范性。

业务服务是目前平台当中基本的服务支持条款。而平台的业务服务支持相较于以往的方式中,无需根据数据库当中的复杂逻辑,只需要提供类似的数据库触发器就可以实现业务逻辑编写。而且开发业务服务程序开发方式及其简单,根据 JAVA 语言配合平台提供的函数资源就可以进行简单的业务逻辑编写。而且这种编写模式可以使得各类软件在进行使用中所提供的业务支持具备一定的一致性。

在界面服务当中,此平台已经给予了完整的版面设计,并且用户可以根据自身的需求进行特殊的页面操作和更改,而且限制条件极少,编写方式简单,使得此平台的设计具备很高的扩展性。平台的界面服务设计依靠的是 Ajax 界面技术,这种技术生成的页面用户可以随意更改页面外观以及页面中连接函数的重新组合搭载编写。这种服务支持以及创新向提出了强调用户体验以及个性化更改的支持,加强了当前的软件使用方便性和简易性,对于应用开发上做出了极大地便利要求<sup>[5]</sup>。

### 2.3 平台的结构设计

在分布式结构下设计的软件数据整合平台大体结构图如图 1。在这个平台结构设计当中主要分为了系统构建部分、应用系统部分、扩展部分。而这些开发部分都依靠网络服务支持。在系统构建的部分当中,系统中存在着五个基本的系统结构。其中有图形化系统构建、页面编辑器、JS 编辑器、样式表编辑器、组件编辑器。在这当中图形化系统编辑器主要是对于整体系统的操作结构以及系统当中的资源数据进行图形化展示。在平台所开发的软件当中,实现图形直观表示以使得系统在可观性上有很大的提高。而页面编辑器则作为基本,是系统操作过程中最为基础的部分。平台所开发的软件系统都是在浏览器页面中进行操作,用页面编辑功能生成整体软件的基本框架。而 JS 与样式表编辑器则主要是进行管理办法提供支持,保证对于平台所开发的软件进行特定分类的标记管理。组件编辑器则根据在数据库当中的资源信息进行整合挑选,符合所开发的软件所需要的相关组件支持,并且对这些组件进行编辑写入,完成整体的软件开发流程。

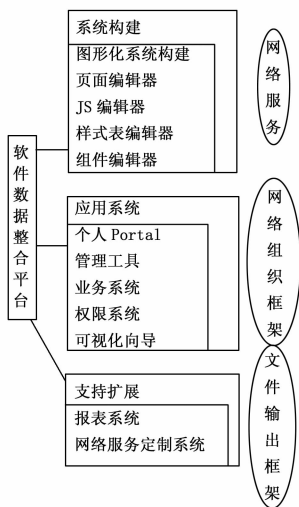


图 1 软件数据整合平台开发结构

在应用系统当中,主要体现在了平台的应用方面上。其中包括了个人 Portal、管理工具、业务系统、权限系统以及可视化向导。个人 Portal 则是用户的登录指令和相关的用户信息,管理工具则是在平台当中对于各个所开发的软件提供相应的管理办法和操作,业务系统和权限系统针对企业对于平台的使用时所提供相应的服务支持和规范使用权限,可视化向导则是强调了在进行使用中给予用户一定的指引,帮助用户充分了解平台内部所储存和提供的软件操作方式。

在支持扩展的部分当中有报表系统以及网络服务定制系统。这些部分则主要是面对企业等使用客户方向。根据客户的

相关要求和特定的使用需求,对平台中的软件给予客户相应的服务支持以及软件更新等操作。

在这一套软件系统的开发过程中其不需要任何一个第三方软件,主要根据此平台的界面操作就可以对于企业所提出的要求进行相关的软件设计,说明此平台具备很高的完整性。并且此平台开发具备了诸多特点。

首先此平台的设计是在网络的基础上进行开发设计的,为此对于平台软件的开发设计的局限性极小,只需要在具备浏览器的环境下就可以进行相应的平台设计以及平台中的软件开发功能。此特点可以总结为在任何具备互联网以及浏览器的条件下就可以工作,对于相应的软件下载都是不需要的。使用上极其方便。其次此系统的占用内存极少,对于开发过程的主机配置要求不高,在有 512 兆的内存主机环境下就可以开发。而后是平台的开发模式采用的是模块化的开发模式,这种模式存在着极高的扩展性和可调节性,对于软件和平台的开发限制得到了解放。而且平台设立了图形化系统搭建,这种模式决定这对于系统界面的修改可以依靠图形界面,并且依靠拖拽的方式就可以更改系统的模块结构。而且页面所采用的静态模式减少了服务器的解压缩压力,加快了整体系统的使用速度和加载使用速度。并且这个平台可以进行团体开发模式以及报表设计器等多功能工具。开发者和用户可以同时在页面上进行操作,而不影响到整体系统平台的稳定性。

### 3 实际应用情况及效果

根据以上的理念设计出一款用于商用软件的分布式软件数据整合平台。根据某公司要进行新型软件更替的要求,依靠此平台根据该公司的实际应用需求进行软件开发设计。总共开发出了智能机房巡检系统、库房储存管理系统、新闻信息发布系统、人力资源管理系统、财务及固定资产管理系统、信息资源调配系统、公司内部安全管理系统、市场信息资源分析系统这八大系统。并且这些系统都实现了单点登录以及身份认证的要求,以及面对系统所需要服务和管理的部分的需求。在实际应用情况中,系统的使用情况客观,而且操作简便具有很高的实用性。并且从开发统计来看,所有系统的开发周期都缩短了将近 50 天,整体系统的开发效率有很大的提高,大约提高了 40%,在开发成本上节约企业用于软件开发以及后期维护和部分硬件投资大约 55%。从企业经济效益上来看,对于企业的整体日常维护起到了极高的加强以及资金的投入。而在企业现代化发展来看,大大的提高了企业的信息化、现代化资源整合的发展,有着明显的提高。

### 4 结论

目前信息化发展以及大数据时代的到来,对于当前的软件开发以及各行业中的应用软件管理提高了很大的要求。本文根据当前的社会现状并且根据商用软件开发为例,提出一种在分布式计算机系统下的软件数据整合平台,根据此平台的功能、服务、结构设计为思路,做到在商业中各企业用于新型软件开发和管理系统的构建,以提供一个可选软件平台。通过实际的应用成果来看,在此平台设计的软件具备更高的开发效率和节约成本,并且使用方便有很高的可行性。

#### 参考文献:

[1] 周小明,金成明. 辽宁电力软件资源整合平台设计与开发 [J].

电力信息化, 2012, 10 (2): 65-68.

- [2] 朱碧琴. 不动产登记信息数据整合及管理基础平台建设 [J]. 低碳世界, 2016, 5 (13): 145-202.
- [3] 王春海, 刘晓辉, 白凤涛. 信息资源整合中的风险及其防范策略 [J]. 电力信息化, 2009, 7 (9): 43-48.

工程中, 由于程序结构或者逻辑都比较复杂, 获得程序的测试用例、覆盖信息、执行结果等信息需要耗费大量的时间, 会降低测试效率。这些不足在后续研究中会继续优化改进。

数据来源, 西门子套件集是由 Software-artifact Infrastructure Repository (SIR)<sup>[16]</sup> 提供, 大部分由 C 语言编写, 每个程序都有正确版本、错误版本和测试用例, 并且错误都是由人工添加进程序的, 有语句赋值错误、谓词错误等, 错误类型都和实际工程中可能发生的类型相似。但是, 并不是套件集所有的错误版本都适合本次实验, 例如 print\_tokens 的错误版本 4 的错误语句存在于头文件中, 不能统计错误语句的覆盖信息; 例如 replace 的错误版本 27 和错误版本 32, 在执行测试用例过程中会发生错误, 从而会产生程序异常终止, 不能得出输出结果数据。

为了检验算法的有效性, 引入定位有效率作为评价指标, 算法的定位有效率是指在定位错误过程中所花的精力, 也就是检测出错误百分比同已检测代码百分比的比值, 假设 A 算法和 B 算法同样检测出了 55% 的错误, 但是 A 算法已检测代码占总代码的 30%, 而 B 算法已检测代码占同代码的 50%, 那么显然, A 算法的定位有效率的值小于 B 算法的定位有效率的值, A 算法的计算效率比 B 算法的计算效率更高, 性能更优。图 3 展示了 EGA-BPN 同 GA-BPN 和 BPN 的定位有效率的比较。

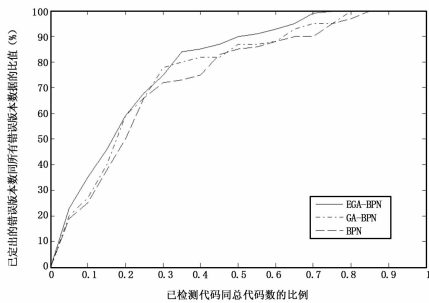


图 3 EGA-BPN、GA-BPN、BPN 的定位有效率比较

根据图 3 数据, 我们可以看出, 在已检测同样的代码数的情况下, EGA-BPN 能定位更多的错误, 同样, 在定位同样多错误的情况下, EGA-BPN 只需检测更少量的代码, 相比较而言, EGA-BPN 具有更高的效率和精确度。

### 4 总结与展望

本文将遗传算法、BP 神经网络和正交试验设计融合, 应用于错误定位领域, 提出了增强 GA-BP 神经网络的软件错误定位算法, 实验证明, 该算法在定位错误的效率和精确度方面确实比 GA-BPN 和 BPN 具有一定的提高。虽然该错误定位算法取得了良好的实验结果, 但是该算法还存在很多不足, 例如: 实验结果只是给出了错误语句的可疑度排名, 还需要工作人员逐一排查, 手动修改错误并再次执行程序以确认错误是否已修改; 另外, 神经网络预测的准确性同训练数据有较大关系, 训练数据越多, 预测结果相对来说就更准确, 但是在实际

- [4] 高磊, 肖建飞, 范俊杰, 等. 企业门户与目录系统的研究与应用 [J]. 电气信息化, 2011, 9 (2): 138-143.
- [5] 李立博. 面向服务的多源异构数据整合平台的设计 [J]. 计算机工程与设计, 2011, 32 (1): 141-144.

### 参考文献:

- [1] Sterling C D, Olsson R A. Automated bug isolation via program chipping [J]. Software: Practice and Experience, 2007, 37 (10): 1061-1086.
- [2] Zhang X, Gupta N, Gupta R. Locating faulty code by multiple points slicing [J]. Software: Practice and Experience, 2007, 37 (9): 935-961.
- [3] Wong W E, Qi Y. Effective program debugging based on execution slices and inter-block data dependency [J]. Journal of Systems and Software, 2006, 79 (7): 891-903.
- [4] Abreu R, Zoetewij P, Van Gemund A J C. On the accuracy of spectrum-based fault localization [C]. Testing: Academic and Industrial Conference Practice and Research Techniques Mutation, 2007: 89-98.
- [5] Jones J A, Harrold M J, Stasko J. Visualization of test information to assist fault localization [C]. Proceedings of the 24th international conference on Software engineering, ACM, 2002: 467-477.
- [6] Masri W. Fault localization based on information flow coverage [J]. Software Testing: Verification and Reliability, 2010, 20 (2): 121-147.
- [7] Murtaza S S, Madhavji N, Gittens M, et al. Diagnosing new faults using mutants and prior faults (NIER track) [C]. IEEE 33rd International Conference on Software Engineering, 2011: 960-963.
- [8] Yu K, Lin M, Gao Q, et al. Locating faults using multiple spectra-specific models [C]. Proceedings of the 2011 ACM Symposium on Applied Computing, 2011: 1404-1410.
- [9] Zhang X, Gupta N, Gupta T. Locating Faults through Automated Predicate Switching [C]. In the 28th International Conference on Software Engineering (ICSE. 06), May 2006: 272-281.
- [10] Cleve H, Zeller A. Locating Causes of Program Failures [C]. In the 27th International Conference on Software Engineering (ICSE. 05), 2005: 342-351.
- [11] Wong W E, Qi Y. BP neural network-based effective fault localization [J]. International Journal of Software Engineering and Knowledge Engineering, 2009, 19 (4): 573-597.
- [12] 史忠值. 神经网络 [M]. 北京: 高等教育出版社, 2009.
- [13] 张柯, 张德平, 汪帅. 基于增强径向基函数神经网络的错误定位方法 [J]. 计算机应用研究, 2015, 3.
- [14] Eric Wong W, Vidroha Debroy and et al. The DStar Method for Effective Software Fault Localization [J]. IEEE Transactions on Reliability, 2014, 63 (1): 290-308.
- [15] Wong W E, Debroy V, Thuraisingham B and et al. RBF Neural Network-based fault location [J]. Technical report UTDCS-20-10, 2010.
- [16] SIR, <http://sir.unl.edu/portal/index/php> [EB/OL].