

基于云环境下的科研管理服务平台建设研究

李颖

(安徽省科学技术情报研究所, 合肥 230091)

摘要: 以云环境下的科研管理服务平台建设为研究对象, 通过云计算的体系架构搭建具有分布式存储、虚拟化应用的科研管理服务平台, 平台用户包括管理员、科研管理服务工作者和科研人员三种用户, 平台提供的功能主要包括用户管理、科研管理、科研服务管理、系统管理; 服务平台利用 HDFS 分布式文件系统功能对海量的数据进行处理, 通过实践的应用证明了该平台对科研课题的申报和科研统计具有很好的服务能力, 极大地提高科研管理服务的效率, 推动科研项目成果的转化规范化发展。

关键词: 云计算; 科研管理; 科研申报; 信息统计

Research on the Construction of Scientific Research Management Service Platform Based on Cloud Environment

Li Ying

(Anhui Institute of Science and Technology Information, Hefei 230091, China)

Abstract: Taking the construction of scientific research management service platform in cloud environment as the research object, a scientific research management service platform with distributed storage and virtualization application is built through the architecture of cloud computing. The users of the platform include administrators, scientific research management service workers and scientific researchers. The functions provided by the platform mainly include user management, scientific research management, scientific research service management and system. Administration. The service platform uses the function of HDFS distributed file system to process massive data. Practical application proves that the platform has a good service ability for the declaration of scientific research topics and scientific research statistics, greatly improves the efficiency of scientific research management services, and promotes the transformation and standardization of scientific research projects.

Keywords: cloud computing; scientific research management; scientific research declaration; Information statistics

0 引言

科研管理服务是技术革新创新发展的后勤保障服务之一, 随着我国市场经济改革不断深化, 自主创新已经成为时代发展的重要支撑, 科研项目的数量也逐年增长, 经费数额不断的增大, 科研成果越来越多样化, 这就给科研管理服务提出了新的挑战和要求。科研管理服务信息化是提高科研管理水平的重要手段之一, 在面对海量科研项目信息和复杂的科研内容方面, 需要一种可靠的、安全的服务平台进行高效的管理, 而云环境则是一种以分布式处理虚拟化服务的一种资源服务技术, 可以大大提高科研管理服务信息化的质量和效率, 为此构建云环境下的科研管理服务平台具有非常重要的价值。本文以某科研单位科研管理工作内容为出发点, 通过分析科研管理信息化需求进行科研管理服务平台的设计, 划分用户管理、科研管理、科研服务管理和系统管理四大模块, 对功能和数据库进行设计, 搭建应用环境进行系统测试, 通过应用得到反馈, 平台对

于该科研单位科研管理服务信息化具有很好的应用效果。

1 关键技术

1.1 云计算

云计算中的“云”表示具有相当规模, 云计算也就相当规模的计算。在互联网发展的进程中, 本地计算已经无法满足海量数据的计算, 为此通过一种基于网络的虚拟化资源进行按需分配形式的计算, 动态扩展资源满足不同需求的用户对信息资源的利用。云计算的运算能力非常的强大, 甚至可以达到每秒 10 万次, 这种强大的计算力在信息爆炸的社会里具有非常重要的实用价值。

云计算是一种利用散落在各地的廉价计算机构建的大规模计算和存储服务云环境, 它可以按照用户的需求获取网络应用资源, 其服务模式主要包括软件即服务、平台即服务、基础设施即服务和数据即服务等^[1]。云计算可以动态的扩展和伸缩, 支持数据的写入与读取和软件的远程应用, 云计算采用虚拟化技术对资源进行整合可以保证数据的安全性和容错能力。云计算是一种按需使用并按使用量付费的网络访问模式, 用户可以通过网络使用存储、软件应用、服务器等计算机资源, 减少本地资源的利用, 降低成本。

收稿日期: 2019-06-12; 修回日期: 2019-07-04。

基金项目: 合肥市软科学研究项目重大课题(2015FFCZ5498)。

作者简介: 李颖(1984-), 女, 安徽合肥人, 硕士, 助理研究员, 主要从事科技信息管理, 科技情报方向的研究。

1.2 HDFS 分布式存储

HDFS 是 Hadoop 框架下的基础分布式存储架构，它是由 Google 公司提供的开源平台，可以将大量分布在全世界各地的廉价计算机通过网络连接在一起，每一计算机都可以作为一个 DataNode 节点，由服务器 NameNode 进行统一的管理和资源的分配^[2]。HDFS 系统架构如图 1 所示。

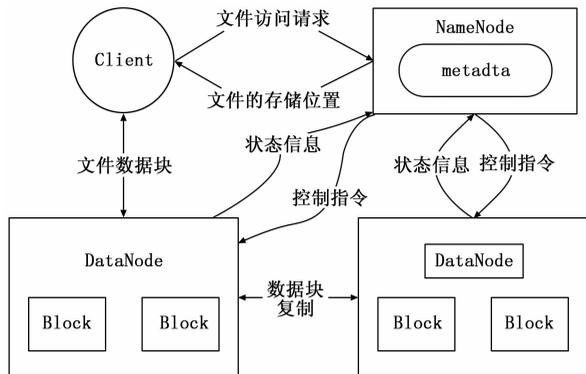


图 1 HDFS 系统架构

在 HDFS 分布式文件系统中，有且仅有一个主服务器命名为 NameNode。NameNode 服务器节点可以对集群内所有 DataNode 节点上的文件块和元数据进行统一的协调和管理。管理内容包括定位元数据位置、调度数据存储的节点位置、查询元数据等。

在 NameNode 服务器节点中，FSImage 是保存 HDFS 文件系统目录和文件的检查点，每次元数据的写入都会使命名空间发生变化，而 Editlog 事务日志记录了分布式文件系统的每一次操作信息，能够有效的避免因多次操作而使 FSImage 文件频繁被更改。

NameNode 节点能够监听来自所有 DataNode 节点和客户端的发生事件，对 DataNode 节点和客户端的操作请求进行处理，其中节点操作请求信息主要包括数据块状态信息、节点存储出错信息、心跳响应等。

心跳检测是 DataNode 节点通过发送心跳信息来周期性地向 NameNode 节点汇报自己当前的节点状态，若 NameNode 超过固定时间还未接收到 DataNode 心跳信息，则诊断为该 DataNode 节点出现故障。

1.3 虚拟化技术

虚拟化技术将计算机的软硬件以一种虚拟化的方式提供给用户使用，一方面降低用户客户端数据处理及软件应用的开销，另一方面对资源进行统一管理提高资源的安全性^[3]。虚拟化技术可以解决物理硬件老旧更迭造成大量的资源浪费问题，以产能低的物理设备重组，通过网络将底层物理硬件资源进行分区，为用户提供远程的多样化服务。采用虚拟化技术构建虚拟机，虚拟化层是一个中间件，它对底层的硬件资源如硬盘、CPU、内存等进行封装，为用户层提供多份抽象资源。

在云环境下，虚拟化技术可以将底层硬件与应用服务

相隔离，将单个资源分裂出多个虚拟资源同时面向多个用户提供服务，同时也可以将多个资源聚合成一个虚拟资源进行管理。

2 系统设计

2.1 总体设计

科研管理服务平台是对科研信息进行收集、整理、组织、存储、加工、使用与维护的综合性服务平台^[4]，主要功能模块划分为科研信息管理模块、科研信息统计模块、科研信息查询模块以及系统管理模块，各个模块下结合科研管理服务的需求由下设了多个子功能，用户管理模块包括：管理员用户管理、科研人员管理和科研服务人员管理；科研管理模块包括：科研类型管理、科研信息管理、科研文件管理；科研服务管理模块包括：科研信息统计与查询、科研成果统计与查询和科研经费统计与查询；系统管理模块包括：登录日志管理功能、数据管理功能。科研管理服务平台总体结构如图 2 所示。

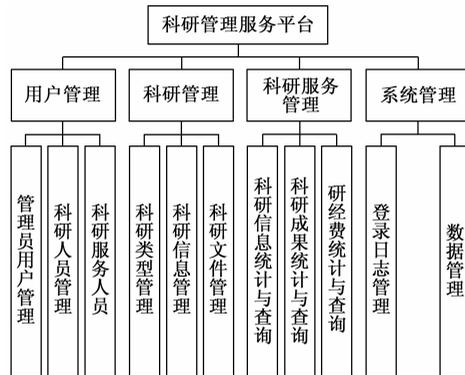


图 2 科研管理服务平台总体结构

2.2 功能模块设计

2.2.1 用户管理功能设计

科研管理服务平台用户包括管理员、科研人员和科研服务人员，不同用户使用平台拥有不同的权限，管理员可以进行系统功能的管理和系统安全的管理，包括对系统用户的添加与编辑，查看系统登录日志信息和进行数据的备份与恢复；科研人员身份登录系统可以进行科研信息的申报和科研文件的上传与下载；科研服务人员登录系统可以进行科研信息、科研成果、科研经费的统计与查询。

所有用户的用户名和初始密码均由管理员提供，用户使用用户名、密码登录系统时系统会根据用户名及密码的正确性进行判断，如果错误则返回登录界面，并提示“用户名或密码错误”，如果用户名及密码正确则判断用户的身份，对应进入用户权限范围内的操作界面

2.2.2 科研管理功能设计

科研管理功能是针对科研人员设计的功能模块，科研类型管理中可以针对科研人员的科研项目属性添加类型，再将科研信息添加到指定的科研类型下进行申报，科研文

件的上传与下载要根据科研信息进行匹配上传，将科研信息与科研文件对应。

科研管理功能的流程是科研人员首先在科研类型管理中选择所要添加科研信息所属的类型，如果没有已经设置好的类型，则需要添加新的科研类型，之后再 将科研信息放置在指定的科研类型中，科研文件的上传要指定科研信息，将与科研信息相匹配的文件上传到指定的目录中，当科研人员需要下载科研文件时只需要进入科研信息管理中 找到对应的文件下载即可。

2.2.3 科研服务管理功能设计

科研服务管理功能是为科研服务人员提供的功能模块，科研服务人员可以对平台上所有科研项目的信息进行统计，查询指定的科研项目内容。同样还可以对已经转化为科研成果的项目进行统计与查询，对各个科研项目的经费进行统计与查询。

以科研服务人员身份登录系统后可以按照科研项目的日期进行统计查询，也可以按照项目的类型、项目的申报人等进行统计与查询，查询结果可以生成 excel 表格下载到本地计算机上。

2.2.4 系统管理功能设计

系统管理功能只能由管理员进行操作，管理员身份登录系统后可以查看每日的系统登录日志信息，对应显示登录的 IP、用户名等信息，便于管理员对非法登录用户进行跟踪。为了保证科研管理服务平台数据的安全性，管理员需要定期进行系统数据的备份，如果发生意外事故可以尽可能的将损失降到最低。

2.3 数据库设计

科研管理服务平台数据的写入与读取建立在云环境应用之下，采用 HDFS 分布式文件系统进行数据处理，并通过 MapReduce 编程模型进行数据的处理^[5]。数据处理流程如图 3 所示。

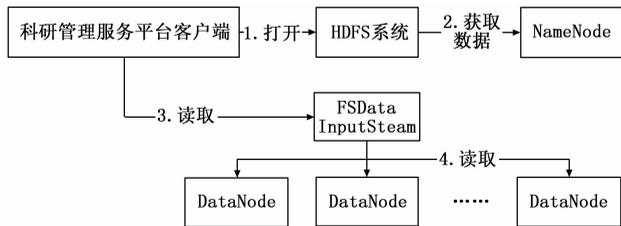


图 3 数据处理流程

科研管理服务平台数据库的设计根据功能模块的划分和数据的类型建立数据库表，主要包括：用户管理信息表、用户信息表、科研信息表等。

用户管理信息表字段主要包括：用户名称、用户密码、用户权限。用户管理信息表如表 1 所示。

用户信息表字段主要包括：用户 ID、用户姓名、用户性别、身份证号、单位名称、联系电话、地址等。用户信息表如表 2 所示。

表 1 用户管理信息表

字段名称	字段类型	字段长度	是否为空
用户名	char	10	NO
用户密码	char	10	NO
用户权限	char	3	NO

表 2 用户信息表字段

字段名称	字段类型	字段长度	是否为空
用户 ID	char	10	NO
用户姓名	char	10	NO
用户性别	char	2	NO
身份证号	char	18	NO
单位名称	char	20	NO
联系电话	char	11	NO
地址	char	50	NO

科研信息表字段主要包括：项目编号、项目名称、申报人、项目类型、科研成果类型、发布日期、科研内容、备注等。科研信息表如表 3 所示。

表 3 科研信息表

字段名称	字段类型	字段长度	是否为空
项目编号	char	10	NO
项目名称	char	30	NO
申报人	char	10	NO
项目类型	char	10	NO
科研成果类型	char	10	NO
发布日期	Datetime	8	NO
科研内容	Varchar	100	NO
备注	Varchar	100	YES

3 系统应用分析

3.1 环境搭建

此次科研管理服务平台应用环境的搭建采用了 6 台普通计算机进行云环境的搭建，其中一台作为 NameNode 服务器，5 台作为 DataNode 节点，使用虚拟机 vmware 在 Linux 操作系统中安装 Java 开发环境。

硬件环境 NameNode CPU Xeon W3550、DataNode CPU Intel core2 Duo E6600、各节点硬盘 20G、各节点内存 4G。软件环境操作系统 ubuntu14.04.3、开发工具 Myeclipse2014、Hadoop 版本 2.20、JDK 版本 1.7 集群架构如图 4 所示：

Hadoop2.20 提供了两个 NameNode 节点，一个作为 active 节点，另一个作为 standby 节点。通过 zookeeper 监控 active 节点状态，二者只有一个 NameNode 节点管理集群工作，另外一个节点作为备用。当 active 节点出现故障则 zookeeper 唤醒 standby 节点作为 active 节点。在两个集群中 zookeeper 集群提供元数据的同步与 NameNode 的切换。

HDFS 集群作为文件存系统调用 Hadoop java api 进行

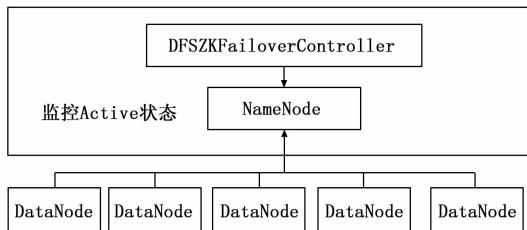


图 4 集群架构

接口操作实现数据的访问。在 hdfs-site.xml 配置文件中应用程序使用数据库的大小和文件数，core-site.xml 文件设定 NameNode 的端口号及 IP 地址。集群节点部署如表 4 所示：

表 4 集群节点部署

主机名	IP 地址	运行进程
NameNode	192.168.1.101	NameNode
DateNode	192.168.1.102	ResourceManager
DateNode	192.168.1.103	DataNode NodeManager
DateNode	192.168.1.104	DataNode NodeManager
DateNode	192.168.1.105	DataNode NodeManager
DateNode	192.168.1.106	DataNode NodeManager

平台的整体架构中除了云环境的云计算服务器、数据存储服务器外，还包括 web 服务器、负载均衡设备^[6]。云计算服务器进行科研管理信息资源的统一管理和计算；数据存储服务器进行数据和文件的存储；web 服务器提供平台的访问；负载均衡设备动态的进行资源分配。

3.2 系统测试

科研管理服务平台开发完成后首先对系统各项功能测试，以保证系统的完整性、可靠性和实用性。系统功能测试内容及结果如表 5 所示。

表 5 系统功能测试内容及结果

测试内容	测试用例	测试结果
用户登录测试	正确的用户名及密码	登录成功
	错误的用户名或密码	登录失败
	不同权限用户名及密码	对应进入用户权限范围功能界面
信息管理测试	添加科研分类	提示“科研分类添加成功”
	添加科研信息	提示“添加科研信息成功”
	删除科研信息	提示“删除科研信息成功”
文件上传与下载测试	上传文件	提示“文件上传成功”
	下载文件	提示“文件下载成功”
统计与查询测试	统计指定日期范围的科研项目	显示统计内容
	查询指定名称的科研项目	显示指定名称科研项目内容

完整性和可靠性，能够达到软件的使用需求效果。

3.3 系统应用评价

将该系统应用于某科研单位进行科研管理服务应用，按照该科研单位的科研管理流程使用此系统进行信息化管理，使用 3 个月的时间进行效果的评价，用户均对此系统表示出了认可，可以满足科研管理服务的基本需求，能够提高科研管理服务的效率，并且在科研工作中逐步习惯了使用系统进行信息交互，使科研管理服务平台成为科研管理工作不可分割的组成部分。

通过用户的反馈可以看出，系统在操作上具有用户类型划分的明确性，能够简化科研人员项目申报流程和提高科研服务人员服务效率。在系统安全方面采用了数据集中式的管理方法，避免了因个人原因而导致的科研项目信息和科研成果丢失的问题。在科研经费使用情况方面，通过系统可以准确的了解到科研经费落实情况，保证了科研经费支出的可追踪性。

4 结论

1) 科研管理服务是一项信息流非常庞大的工作内容，通过基于云环境下的科研管理服务平台建设可以将科研人员、科研服务人员的工作内容进行有机的结合，提高科研管理服务的效率，降低工作人员服务的压力。

2) 本文根据科研管理服务的工作对象和工作内容设计科研管理服务平台，划分出三种类型的用户和四个主要功能模块，不同类型的用户对应使用权限范围内的功能模块，操作简单，便于快速掌握使用方法。

3) 鉴于科研项目数量的庞大和项目类型的复杂，本文搭建基于 HDFS 的集群云环境进行数据的服务，一方面为数据的写入与读取提供了可靠的服务支持，另一方面可以保证数据的安全性。

4) 经过对系统功能的测试得到的结果是系统功能具有很好的完整性和可靠性，在经过实践应用中获得了广泛的好评，说明了系统的实用性。

参考文献：

[1] 李 松. 开源云计算平台管理支撑平台的研究与实现 [D]. 济南：山东大学，2013.

[2] 张栗棕, 崔 园, 罗光春, 等. 面向大数据分布式存储的动态负载均衡算法 [J]. 计算机科学, 2017, 44 (5): 178-183.

[3] 基于网络功能虚拟化的网络试验平台的设计与实现 [J]. 计算机学报, 2018, 41 (9): 16-18.

[4] 夏 翔, 华 琳, 韩小鹏, 等. 基于科研训练管理平台数据的统计与分析 [J]. 中国信息技术教育, 2018 (7): 107-108.

[5] 杜 江, 张 铮, 张杰鑫, 等. MapReduce 并行编程模型研究综述 [J]. 计算机科学, 2015, 42 (s1): 2635-2642.

[6] 黄 懋. 基于集群的 HDFS 高可用性研究和实现 [D]. 上海：复旦大学，2012.

通过对功能的测试可以看出，系统功能具有较好的完