

# 独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得安徽大学或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：陈德录

签字日期：2011年5月22日

# 学位论文版权使用授权书

本学位论文作者完全了解安徽大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权安徽大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密的学位论文在解密后适用本授权书)

学位论文作者签名：陈德录

导师签名：刘峰

签字日期：2011年5月22日

签字日期：2011年5月28日

学位论文作者毕业去向：

工作单位：

电话：

通讯地址：

邮编：

## 中文摘要

当前,电子政务建设受到各级政府的高度重视,运用信息化技术更好地服务发展、便民利民、改善民生得到一致认同。2009 年度全国税务系统纳税服务工作会议提出,纳税服务与税收征管共同构成税务部门的核心业务。为纳税人提供优质、高效的纳税服务已成为各级税务机关的共识和努力方向。

但是,当前纳税服务体系还存在一些不足,主要表现在,信息化服务内容和方式单一、征纳双方运维工作量大。为适应纳税人多元化、个性化、快捷高效的服务需求,提高资源利用率,建立统一的易扩充、可定制的纳税服务平台已经成为适应经济社会发展、符合纳税人根本利益的必然要求。尤其是建立像银行一样的国家级、信息共享的国地税通用服务平台是十分迫切的问题。

云计算技术的迅速发展无疑给这些问题的解决带来了契机。云计算技术可以将分散各地的基础设施虚拟化,集中到云计算平台之中,根据 SOA 的思想开发出满足不同需求的应用系统,能够快速部署到服务平台上,纳税人和税务部门可以通过多种方式接入服务平台,享受多元化、个性化的服务。

本文对基于云计算的纳税服务平台的实施方案和功能设计进行了描述,阐述了基于开源 Eucalyptus 云平台和 WebLogic Server、MongoDB 实现私有云平台的实践过程,主要包括以下方面:

1. 对纳税服务信息化的现状以及云计算进行了深入的研究。
2. 提出基于云计算的纳税服务平台的总体架构和实施方案。
3. 完成纳税服务系统中的税收征管档案子系统的主要功能设计与开发。
4. 使用 Ubuntu Enterprise Cloud (UES) 搭建私有云平台并部署税收征管档案子系统。
5. 对税收征管档案子系统进行了测试,测试结果表明基于云计算的应用系统在系统响应速度、数据读写速度方面明显占优,尤其是在高并发或大数据量的情况下。

**关键词:** 云计算; 纳税服务平台; 私有云; 非关系型数据库

# Abstract

E-government services have been popular nowadays for all different levels of governments. There has been a consensus among different levels of governments that it is necessary to utilize information technology into better serving social development, benefiting the public, improving people's livelihood. On the National Taxing Conference (2009), one decision has been made clearly that taxing service, tax collection and governance are the core businesses for Tax Authorities. For Tax Authorities, they all know very well that their task is to provide the taxpayers an efficient and high-quality taxing service.

However, there are still some deficiencies in the current system of taxing services, which could mainly be manifested by two aspects: the weakness of information service and the high cost of Taxing for both Tax collector and taxpayer. In order to provide taxpayer a diversified, personalized and efficient service, and to promote resource utilization, the establishment of a unified and expansive taxing service System must be undertaken. The system is not only going along with social and economic development, but also being in line with the fundamental interests of taxpayers. This system is just like National Inter-bank Information System, which is a platform of information for both National Tax Authority and local Tax Authorities.

Even there are some obstacles in the establishment of the system, Cloud Computing Technology is much helpful to solve these problems. Cloud Computing could make scattered infrastructures virtualized and absorb all them onto Cloud Computing Platform. According to SOA, different applied systems could meet the different needs of applications, which can be quickly deployed to the service platform. If done so, both taxpayers and tax authorities can access service platform in many ways to enjoy diversified and personalized services.

This essay describes Cloud Computing-based taxing service platform designed to implement programs and functions, and explains the processes of practices based on Eucalyptus , WebLogic Server and MongoDB .The main contents of this essay are

follows:

1. The status of informatization of tax service and cloud computing studies.
2. Based on cloud computing tax services platform framework and implementation plan.
3. Finishing the main function of tax administration archives subsystem of tax service system design and development.
4. Using Ubuntu Enterprise Cloud (UES) build private Cloud platform and deploy the tax administration archives subsystem.
5. Tax administration archives subsystem test, test results show that the application system based on cloud computing in the system response speed, reading and writing data speed is very good, especially in the cases of high concurrent or large amount of data.

**Keywords:** Cloud Computing; Taxing services platform; Private cloud; NoSQL

# 目 录

摘 要.....	I
ABSTRACT .....	II
目 录.....	IV
第 1 章 绪论.....	1
1.1 选题背景.....	1
1.2 研究意义.....	2
1.3 发展趋势.....	2
1.3.1 国外纳税服务.....	2
1.3.2 国内纳税服务.....	3
1.3.3 云计算的发展.....	4
1.4 研究内容.....	5
1.5 内容体系.....	6
第 2 章 云计算综述.....	7
2.1 基本概念.....	7
2.2 主要特点.....	7
2.3 部署模式.....	9
2.4 架构层次.....	9
2.5 体系结构.....	10
2.6 主流平台.....	11
第 3 章 纳税服务平台研究.....	14
3.1 现有纳税服务体系.....	14
3.1.1 纳税人面对的问题.....	15
3.1.2 税务部门的难题.....	16
3.2 纳税服务云平台.....	16
3.3 云平台实现方案.....	18
3.3.1 实施进程.....	18
3.3.2 风险点.....	20
3.3.3 工作任务.....	21

第 4 章 纳税服务平台设计.....	23
4.1 总体架构.....	23
4.2 需求分析.....	24
4.3 存储设计.....	26
4.4 网络选择.....	26
4.5 功能实现.....	27
4.5.1 开发环境.....	27
4.5.2 关键功能.....	28
第 5 章 纳税服务平台部署.....	33
5.1 部署私有云.....	33
5.2 管理镜像.....	35
5.2.1 下载镜像.....	35
5.2.2 自建镜像.....	35
5.3 管理实例.....	38
5.3.1 euca系列命令.....	39
5.3.2 HybridFox应用.....	39
5.4 部署应用.....	43
5.5 性能测试.....	44
5.5.1 响应速度测试.....	44
5.5.2 数据读取测试.....	47
第 6 章 结论与展望.....	48
6.1 总结.....	48
6.2 展望.....	48
参考文献.....	50
致 谢.....	54
攻读学位期间发表的学术论文目录.....	55

# 第一章 绪论

近年来,各级政府十分重视电子政务建设,努力打造法治型、数字型、智能型、服务型政府,更好地服务发展、便民利民、改善民生。2009 年度全国税务系统纳税服务工作会议提出,纳税服务与税收征管共同构成税务部门的核心业务。当前,各级国税部门正在金税工程一期、二期的基础上,按照“始于纳税人需求、基于纳税人满意、终于纳税人遵从”的要求,以门户网站、“12366”服务热线、手机短信平台、网上办税平台等多种方式尽力为纳税人提供税法咨询、事项提醒、业务办理等便捷服务,纳税人的满意度逐年提高。

## 1.1 选题背景

国税部门现有的纳税服务主要包括事项提醒、税法咨询、网上办税、税法宣传等。提供纳税服务的应用系统涉及 Windows、UNIX 两大操作系统平台, Microsoft SQL Server、Oracle、Access 等数据库。各个应用系统都是基于特定的业务需求设计和开发,系统之间的互联、互通、互动性较差,数据共享存在困难,普遍存在“信息孤岛”现象。因此,当前纳税服务体系不可避免的存在一些不足。主要表现在,由于征管业务系统与手机短信群发系统不能数据共享,部分纳税人会收到与自己无关的催办业务的短信;纳税人不能定制个性化需求,有时想要的服务无法反馈给国税部门,有时国税部门在充分调研基础上提供的服务也许无法满足一部分纳税人的需要;国税部门提供的应用系统大都需要下载、安装、升级、数据备份等处理,给纳税人带来便捷的同时也带来技术处理方面的负担;纳税人在门户网站中进入不同的应用系统办理涉税业务时需要多次输入用户名和口令;业务数据省级集中方式会导致数据量一般每年增长 200GB 左右,与之相关的系统响应速度正逐渐下降。

国税部门面临的这些问题,其实在政府信息化过程中都已经出现或即将出现。随着业务流程再造和提高工作效率的需要,大的应用系统及其数据逐步集中到省级,造成下级单位原购置的部分小型机或 PC 服务器闲置,也增加了省级

信息中心数据存储、机房扩容带来风险和压力。现有的纳税服务体系建设已经难以适应纳税人多元化、个性化、快捷高效的服务需求。因此，建立具有统一平台的免填单、可定制、易部署、快响应的纳税服务平台已经成为适应经济社会发展、符合纳税人利益的迫切要求。

## 1.2 研究意义

在国税部门建立基于云计算的纳税服务平台，目的是用更少的成本、更快的效率部署应用系统、提升服务质量、丰富服务内容、提高响应速度。对将来实现统一的业务平台建设和国家级数据计算、存储、灾备有一定的借鉴意义。

具体意义包括：

1. 实现现有资源高效利用。以某省国税系统为例，国税部门核心业务系统数据集中到省局，每个市级国税部门都闲置一台小型机，全省闲置近 20 台，还有部分 PC 服务器。而云计算能够将这些基础设施以服务的方式整合利用，可以节约硬件、软件的采购、运行成本，大大减轻省级数据中心的扩容压力。

2. 实现易扩充、可定制、可监管的服务。云计算平台可为纳税人（包括大中小企业、事业单位、个体工商户）和税务部门提供全方位、便捷的服务。税务部门提供的应用系统构建在以 SOA 为基础架构上，能有效利用 SOA 的灵活性，以服务的形式快速部署到云计算平台，并能实现统一监管，规范作业流程，提高系统效率。而纳税人也能够根据自身涉税业务特点按需定制应用系统等服务。

3. 对海量税收数据处理的借鉴意义。随着金税工程三期的实施，海量的税收数据需要实现“国家级”实时处理和查询，在现有应用系统和技术下实现较为困难。通过纳税服务平台实现为建立统一的税收业务平台积累经验。

## 1.3 发展趋势

### 1.3.1 国外纳税服务

国外的纳税服务是伴随着新公共管理运动的兴起而发展的<sup>[1]</sup>。近年来，为适应新经济的发展，把纳税人视为顾客、优化纳税服务已成为各国税收征管改革的发展方向。美国设立小企业管理局，专门负责对中小企业进行服务和扶持。法国成立税务管理中心协助中小企业进行管理，帮助指导建立完整的会计账目。



世界各国政府积极推行电子税务平台，实行税收信息化，甚至已实现全国统一的税收征管服务平台。1994 年美国正式实施“纳税服务现代化”项目，大规模调整和改善税务征管计算机系统，为社会提供更加完善的纳税服务<sup>[2]</sup>。

德国已经形成便捷、高效、统一的税收信息化管理模式，将税务征管的基本业务全部通过网络与信息技术进行集成处理。税务部门在互联网上进行税款征收、档案管理、信息储存、信息交流、税源控管等各种税收管理，为纳税人提供透明、规范的管理和服务，纳税人可以通过互联网方便、快捷地办理相关的涉税事宜。（1）利用计算机给所有开始生产经营活动或者取得收入的单位和个人建纳税人编码和税卡，纳税人通过编码进行纳税申报和缴纳税款；（2）在各州设立税务信息处理中心，州以下的各市、县税务局设立服务信息站，依托计算机对纳税人的纳税申报进行审核、征收税款、缴入国库，州税务信息处理中心和服务信息站通过网络传输信息，从而实现全过程计算机控制；（3）会计（税务）师通过数据信息处理中心及其分支机构为纳税人代理纳税申报业务，该中心已成功地与各州税务信息处理中心实现联网；（4）利用计算机网络对税源实施监控和稽查，计算机网络的应用已发展到了几乎所有基层单位，税收信息化覆盖面非常广泛。

新加坡国内收入局建立了一个统一的有文件图像和工作流程的一体化数据库平台，其中包括“顾客交流”系统、纳税申报系统以及国内收入局内各部门和工作人员之间的交流平台。该平台可以实现自动化处理纳税人申报等税收业务，可以自动创建工作项目实现日常办公自动化。同时，这一平台还能按照已确定的规则，根据税务官员的技能和经验，将不同的工作项目分给不同的人员，实现内部交流的自动化。韩国引入全面计算机化的税务一体化信息系统，运行“网上报税中心”为征纳双方服务。

### 1.3.2 国内纳税服务

1996 年的全国税收征管改革会议确立了“以申报纳税和优化服务为基础，以计算机网络为依托，集中征收，重点稽查”的税收征管模式，首次确定了“优化服务”在税收征管中的基础地位，国务院办公厅批转的税务工作意见又进一步提出和明确了优化服务的问题。2001 年，全国人大常委会通过了新的《税收

征收管理法》。新的《税收征管法》第一次把“纳税服务”提到法律的层面上来，用法律的形式确定了纳税服务作为行政行为的根本性质和法律地位，明确了纳税服务的内涵、内容，纳税服务真正成为各级税务机关的一项重要日常工作。2007年5月，国家税务总局在北京召开全国首次纳税服务工作会议，提出构建法治、公正、服务、对等的税收征纳关系。

目前，各地税务部门在理念上已经实现从“监督与管理”到“服务与管理”的转变，充分认识到纳税服务离不开信息化，便捷、高效、规范、个性化的服务方式实现服务目标的重要保障。国内纳税服务信息化建设起步晚，但发展很快。大力实施金税工程，已实现以省级（国税部门）或市级（地税）数据集中模式实现税收基本业务的信息化处理。虽然积极努力以门户网站、“12366”热线电话、办税平台等方式提供涉税咨询、网上办税、事项提醒、维权救济等服务，但缺乏统一的纳税服务信息化平台。

### 1.3.3 云计算的发展

自2006年谷歌CEO埃里克·施密特（Eric Schmidt）首次公开提出“云计算”和亚马逊（Amazon）推出弹性计算云（EC2）服务以来，云计算时代正式来临，并形成全球性的研究和开发的热潮。美国硅谷目前已经约有150家企业涉及云计算，新的商业模式层出不穷，公开宣布进入或支持云计算技术开发的业界巨头包括Microsoft、Google、IBM、Amazon、Sun、Netsuite、NetApp、Adobe等。除了企业加快布局、积极进军云计算领域之外，美国政府大力推行云计算计划。2010年5月13日，美国政府宣布要把税收监控和举报的网站放到Amazon的云计算服务平台上。日本内务部和通信监管机构计划建立一个名为“Kasumigaseki Cloud”的大规模的云计算基础设施，将在2015年完工，目标是支持所有政府运作所需的信息系统，巩固政府的所有IT系统到一个单一的云基础设施，以提高运营效率和降低成本。

随着中国电子学会云计算专家委员会的成立和2009、2010年连续两届中国云计算大会的召开，标志着云计算在国内的研究也日新月异。但目前云计算建设尚在雏形阶段，研究大多集中于概念推广<sup>[3]</sup>、云体系结构、云存储、云数据管理、虚拟化<sup>[4][5]</sup>、云安全<sup>[6]</sup>、编程模型<sup>[7]</sup>等方面的学术层面以及试商用层面。也

有少数公司已经推出了基于云计算的产品，如趋势科技、奇虎、瑞星等安全厂商纷纷提出了“云安全”计划，八百客公司推出了基于云计算的在线客户关系系统，中国移动正在尝试基于“Blue Carrier”云计算平台解决移动搜索问题，2009年11月3日，新浪的云计算官方 Blog 宣布 Sina App Engine Alpha 版上线。文献[8]探讨了云计算在图书馆中的应用，文献[9]提出了基于云计算的智能家庭构想。国内北京、广东等地方都正在筹划或实施建立用于科研、公益或商用的云计算平台。

云计算用于纳税服务平台建设方面，目前尚未见到国内有相关研究文献。

### 1.4 研究内容

针对当前国税部门纳税服务平台发展的现状和需求，在云计算技术这种新型 IT 基础设施的交付和使用模式（云计算开源技术）基础上，初步实现国税部门纳税服务平台和管理机制。

#### 1. 主要内容

- (1) 当前国税部门纳税服务体系建设及应用现状、面临的问题。
- (2) 云计算的在国内外的发展趋势。
- (3) 基于云计算的纳税服务平台的构建方案。
- (4) 基于云计算的纳税服务平台的系统设计和部分实现
- (5) 基于 Ubuntu Enterprise Cloud (UES) 的纳税服务平台的部署应用。
- (6) 探讨云计算在应用服务开发方面存在的问题，展望省级国税部门全部业务平台建设和实现国家级数据处理的应用前景。

#### 2. 研究范围

从应用结构上来讲，按照从底层硬件到上层应用的顺序，云计算主要由三个层次组成：基础设施即服务（Infrastructure as a Service，简称 IaaS）、平台即服务（Platform as a Service，简称 PaaS）及软件即服务（Software as a Service，简称 SaaS）。本文属于云计算的应用型论文，仅对云计算的平台构建（基础设施即服务）和应用系统开发、部署做重点研究。

#### 3. 研究方法、思路

- (1) 调研分析。以某省国税系统纳税服务体系建设为背景，深入了解服务

体系信息化建设的现状和问题，并对其部分服务对象进行实地走访，掌握纳税人的个性化需求愿景。在调研基础上拟定解决方案。

(2) 理论研究。通过对学术会议材料、理论专著、学术论文等资料的学习和领会，了解云计算的发展前沿技术，并对纳税服务平台建设的可行性进行深入探讨。

(3) 论文写作。结合调研成果、理论学习和项目开发经验，提出省级国税部门基于云计算的纳税服务平台的解决方案，完成课题研究和论文撰写工作。

## 1.5 内容体系

本文分六章详细论述了基于云计算的纳税服务平台的研究与初步实现。

第一章简单阐述了选题背景、项目研究的意义、国内外纳税服务的发展、国内纳税服务平台现状以及云计算的研究现状等内容。

第二章主要介绍了云计算的概念、特点、架构服务，以及常用的部署模式等内容。

第三章对现有纳税服务体系框架和基于云计算的纳税服务平台的实现方案做了阐述。通过对现有纳税服务体系框架的分析，找出存在的问题，并提出了基于云计算的纳税服务平台的实现方案。

第四章在对云计算技术和纳税服务平台业务架构分析的基础上，对纳税服务云平台的业务需求作了分析，并在云计算环境下开发实现了税收征管档案子平台的主要功能。

第五章描述了 Ubuntu 私有云部署过程并以开发实现的子平台为例说明如何在云平台中部署、提供应用系统服务。

第六章为全文总结与研究展望。主要对本文的研究成果加以总结，并展望了基于云计算构建纳税服务平台的美好愿景。

## 第二章 云计算研究

当前, 谷歌 (Google)、亚马逊 (Amazon)、IBM 和微软等信息技术产业的巨头们, 以前所未有的速度和规模在推动着云计算技术和产品的普及。甚至有人预计, 云计算将是未来 3-5 年最值得期待的技术革命。本章将对云计算概念、特点、体系等方面进行简要阐述。

### 2.1 基本概念

云计算<sup>[10]</sup>概念由 Google 首先提出, 目前尚无统一的定义。狭义的云计算是指 IT 基础设施的交付和使用模式, 即通过网络以按需、易扩展的方式获得所需的资源; 广义的云计算是指服务的交付和使用模式, 即通过网络以按需、易扩展的方式获得所需的服务。

云计算 (Cloud Computing) 是网络计算 (Grid Computing)、并行计算 (Parallel Computing)、分布式计算 (Distributed Computing)、效用计算 (Utility Computing)、负载均衡 (Load Balance)、虚拟化 (Virtualization)、网络存储 (Network Storage Technologies) 等网络技术、计算机技术的发展、融合的结果。它以网络为媒介, 能够把多个成本低廉、性能较低的计算单位整合成一个具有强大运算能力的系统, 并借助基础设施即服务、平台即服务、软件即服务等先进的商业模式根据用户需要自动提供强大的运算能力。云计算的核心理念之一就是不断提高“云端”强大运算的能力, 逐步将用户终端上的处理迁移到“云端”, 最终把用户终端变为纯粹的输入输出设备。其实, 相对简单的云计算技术已经在现有的网络服务中随处可见, 例如搜索引擎、网络信箱等, 使用者只要输入简单指令即能得到大量信息。

### 2.2 主要特点

根据现有资料总结和运行情况显示, 云计算具有以下特点:

1. 超大规模。“云”必须具有相当大的规模。Google 的云计算数据计算中

心共有 200 多万台计算机, 目前分布于全球 30 多个不同地点, 这些云设施, 被人们称为“信息时代的核电站”。Amazon、微软、IBM、Yahoo 等的“云”也拥有数十万台或数十万台服务器。企业的私有云一般拥有数百以上服务器。

2. 虚拟化。虚拟化技术将底层的基础设置, 包括服务器、存储与网络设备全面虚拟化, 建立起一个共享的随需而选的运行环境。这个环境支持用户使用各类终端(如 PC、手机等)、在任意地方获取感兴趣的资源和服务。用户无需了解、也无法了解应用服务和所获取的各种资源的具体位置。

3. 高可靠性。云计算使用数据多副本容错、计算中间结果持续化、计算节点同构可互换等技术, 从而保证其性能的可靠性, 任意一个节点或者数个节点同时出现故障, “云”中存储的数据不会丢, 提供的服务不会中断。

4. 通用性。一般来说, 云计算(公共云)不针对特定的用户和应用。在云计算平台的支撑下应用服务供应商或终端用户可以在云基础设施上开发、部署千变万化的应用。例如 Google App Engine、Amazon Web Services 以及创建随需应变的应用程序的 Force.com 平台。

5. 高可扩展性。云计算平台的规模可以动态伸缩, 服务节点的数量在一分钟或几分钟之内能够根据数据流量的增减程度自动地动态增加或减少, 既能满足应用和用户规模增长的需要而又不会造成服务资源浪费。

6. 按需服务。“云”是一个庞大的资源池, 用户可以根据需购买或租赁。例如 Amazon 提供的简单存储服务(S3)和弹性计算云服务(EC2)。

7. 极其廉价。云计算的高可靠性使得云计算平台可以采用极其廉价的节点构成, 通用性使资源的利用率得到大幅提升, 自动化集中式的管理可以大大降低数据中心管理成本。因此, 用户可以充分享受云计算的低成本优势。纽约时报曾使用 Amazon 的弹性计算服务, 在 24 小时之内就把历史上所有的文章从 TIF 格式转化为 PDF 格式, 一共花费 240 美元。如果纽约时报使用自己的机器, 需要数月的时间或更多的费用才能完成这一任务。

8. 潜在的危险性。云计算服务能够提供计算服务和存储服务。对于持有敏感数据的政府机构、商业机构来说, 选择云计算服务应充分考虑数据故意丢失或泄密等潜在危险。

## 2.3 部署模式

一般说来,云计算有三种部署模式,即公共云、私有云和混合云。

公共云(Public cloud)适合中小型企业使用,由第三方建设运营。云计算服务供应商把许多不同用户的作业在云内的服务器、存储系统和其它的基础设施上混合在一起。最终用户不知道运行其作业的一台服务器、网络或磁盘上还有哪些用户,也不知道具体位置。目前,典型的公共云有 Google App Engine、Windows Azure Platform、Amazon 的 AWS、Salesforce.com 以及阿里巴巴等。公共云的最大优点是所应用的程序、服务及相关数据都存放在公共云的提供者处,用户无需做相应的投资和建设。但是,由于数据不是存储在自己的数据中心,安全性存在一定风险,服务的可用性也不受用户控制,这方面也存在一定的不确定性。

私有云(Private clouds)比较适合于有众多分支机构的大型企业或政府部门。由单个用户自己按需提供基础设施,拥有服务器、网络和磁盘,控制应用程序在何处运行,可以决定允许哪些用户使用基础设施。相对于公共云,私有云部署在企业内部,数据安全性、系统可用性都可由自己控制,但是投资较大,尤其是一次性的建设投资较大。

混合云(Hybrid clouds)把公共云模式与私有云模式结合在一起。混合云秉承公共云节省成本的优点,同时可部分实现私有云所能提供的控制和合规。混合云提供根据需要且在外部预配置的扩展规模的承诺,但增加了确定如何在这些不同环境之间分配应用程序的复杂性。

## 2.4 架构层次

根据虚拟化技术在云计算中的应用,云计算的架构层可分为应用程序、服务和基础设施。这些层不仅封装了按需提供的资源,还定义了一个新的应用程序开发模式,如图 2-1。

1. 软件即服务(SaaS)。SaaS 在最高层,其特点是包含一个通过多重租用(Multitenancy)根据需要作为一项服务提供的完整应用程序。所谓“多重租用”是指单个软件实例运行于提供商的基础设施,并基于 Web 前端为多个客户提供服务。

2. 平台即服务 (PaaS)。PaaS 为中间层, 是对开发环境抽象的封装和对有效服务负载的封装, 提供软件和产品开发工具的一种云服务, 这些工具由提供者托管于它们的硬件基础设施。用户可以使用这个平台以及所提供的 API 或是通过一个用于开发的图形用户界面来创建应用程序。PaaS 产品可执行各个阶段的软件开发和测试, 也可专用于某个领域。

3. 基础设施即服务 (IaaS)。IaaS 处在最底层, 是一种作为标准化服务在网上提供基本存储和计算能力的手段。服务器、存储系统、交换机、路由器和其他系统协作 (例如通过虚拟化技术) 处理特定类型的工作负载, 例如批处理或者峰值负载期间的服务器/存储扩大。从某种意义上说, 可以把云计算认为是第二波虚拟化的推广和应用, 云计算把虚拟化进行得更深入、更彻底了。

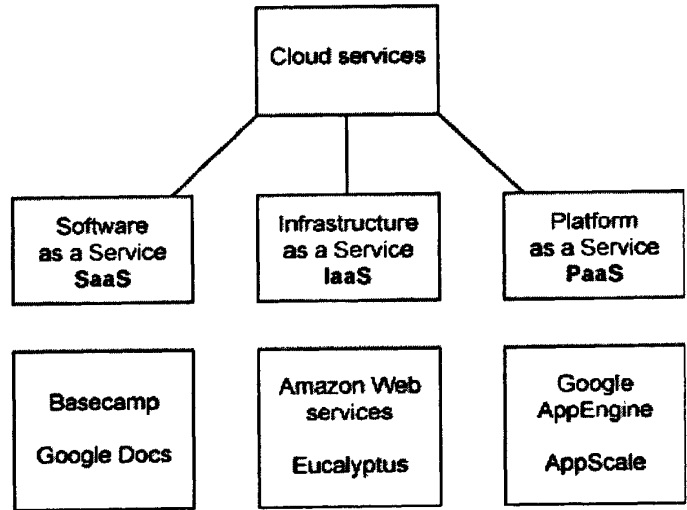


图 2-1 云计算架构层和主要商业示例

Figure 2-1 Cloud computing architecture layer and the main commercial sample

2.5 体系结构

云计算平台连接了大量并发的网络计算和服务, 可利用虚拟化技术扩展每一个服务器的能力, 将各自的资源通过云计算平台结合起来, 提供超级计算和存储能力。通用的云计算体系结构如图 2-2 所示。



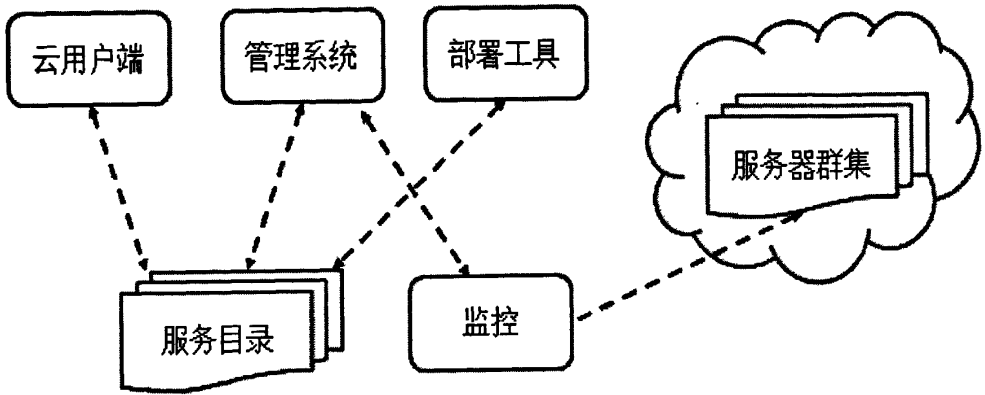


图 2-2 云计算体系结构示意图

Figure 2-2 Schematic diagram of cloud computing architecture

1. 云用户端。云用户端是用户使用云的入口，提供云用户请求服务的交互界面，用户可以通过 Web 浏览器或手机等使用服务。
2. 服务目录。用户在取得相应权限后可以选择服务，也可以退订已有服务，在云用户端界面生成相应的图标或列表来展示已选择服务。
3. 管理系统和部署工具。提供对用户、基础设施、资源和服务进行管理的功能。
4. 监控。监控和计量云计算平台资源的使用情况，以便根据数据流量迅速完成节点同步配置、负载均衡，确保资源能顺利分配给合适的用户。
5. 服务器群集。虚拟的或物理的服务器，由云计算平台控制器负责管理，负责提供计算、Web 应用、存储等服务。

## 2.6 主流平台

目前，全球各大 IT 企业纷纷推出自己的云计算产品和解决方案，云市场中充斥了各种各样的云计算产品与服务，涉及到基础设施层、平台、软件等各层，其中被采用最多的主要是 Google、Amazon、Microsoft、IBM、Yahoo、SalesForce、VMware 等。

### 1. Google 云计算

Google 公司作为云计算的先驱，其云计算技术相对成熟。Google 的产品既

有面向个人消费类产品，如网络搜索、Gmail、Google Docs 等，也有企业级产品 Google App Engine。Google 在 2008 年 4 月推出的 Google App Engine 是 PaaS 应用的典型代表，也是面向开发者的。允许开发人员编写应用程序，然后把应用构建在 Google 的基础架构上。App Engine 方便用户开发新的服务逻辑，并且在云计算平台上运行这些新的服务。Google App Engine 可让用户在其基础架构上运行网络应用程序，App Engine 应用程序易于构建和维护，可根据访问量和数据存储需求的增长弹性扩展。

## 2. Amazon Web Service

Amazon 云计算的优势在于它是应用者而非 IT 系统制定者，云计算平台是开放的，具有很大的灵活性。Amazon Web Service 是一个典型的 IaaS 服务，包括简单存储(Simple Storage Service, S3)、弹性计算(Elastic Compute Cloud, EC2)、消息传递(Simple Queuing Service, SQS)、简单的数据库管理(SimpleDB)等。

## 3. Microsoft Azure 平台

Azure 平台是微软为开发者提供的的一个云开发平台，它是基于 PaaS 模式的。包括了 Windows Azure 操作系统、SQL Azure 数据库和 AppFabric。最底层的是 Windows Azure 这个操作系统。它提供了 Compute(计算)，Storage(存储)，以及 Manage(管理)这三个主要功能以及对用户而言透明的 Fabric。Fabric 包含负载均衡，硬件抽象等功能。一般而言，用户并不需要了解 Fabric 内部是如何工作的，就可以充分利用 Windows Azure 的各种特性。在这之上，还提供了中间件产品 AppFabric，数据库产品 SQL Azure，以及其他一些 building block 产品，例如 Dallas。开发人员可以直接在 Windows Azure 之上进行开发，也可以利用 AppFabric，SQL Azure 等产品的各种特性。

## 4. IBM 蓝云

简单地讲，IBM 蓝云就是一个先进的基础架构管理平台，能够对企业现有的基础架构进行整合，通过自动化技术、虚拟化技术，帮助企业构建私有的云计算中心，从而实现企业软、硬件资源的快速部署、统一管理、自动分配、自动备份，实现应用服务和数据存储对企业硬件资源和软件资源的共享。

### 5. 开源云计算平台

Google、Amazon、Microsoft 等公司的云计算平台都属于公有云，并且需要按需收费。对于大多数企业，尤其是政府部门来说，更希望将应用服务运行在私有云平台之上，保证数据的安全性和可用性。开源云计算平台较好地解决了这一问题。目前，主要的开源云计算平台有 Eucalyptus、AbiCloud、10gen、Enomalism、Nimbu 等。

Eucalyptus (Elastic Utility Computing Architecture for Linking Your Programs To Useful Systems) 是 Amazon EC2 的一个开源实现，它与商业服务接口兼容。和 EC2 一样，最初 Eucalyptus 依赖于 Linux 和 Xen 进行操作系统虚拟化，但是 2010 年 8 月最新发布的 Eucalyptus Enterprise Edition 2.0 已经支持 VMware 以及 Microsoft Windows Guest。

### 第三章 纳税服务平台研究

纳税服务平台的信息化是伴随着税收征管信息化建设而发展的，高度信息化的税收征管一定可以提供高效便捷的税收征管业务方面的服务。但是，纳税服务的内涵不仅仅局限于征管服务，纳税人更需要税务机关提供个性化、多元化、全方位的纳税服务。

#### 3.1 现有纳税服务体系

目前，国税部门的纳税服务体系主要包括由门户网站和“12366”服务热线提供的税法咨询服务、由手机短信平台提供的涉税事项提醒服务、网上办税平台提供的部分业务办理服务、为纳税人提供免费的单机版应用软件服务(如图 3-1)<sup>[11]</sup>。然而，在这种纳税服务体系下，纳税人除了享受信息化服务的便利之外，同时也经常感受到信息化这剂良药的“副作用”困扰。

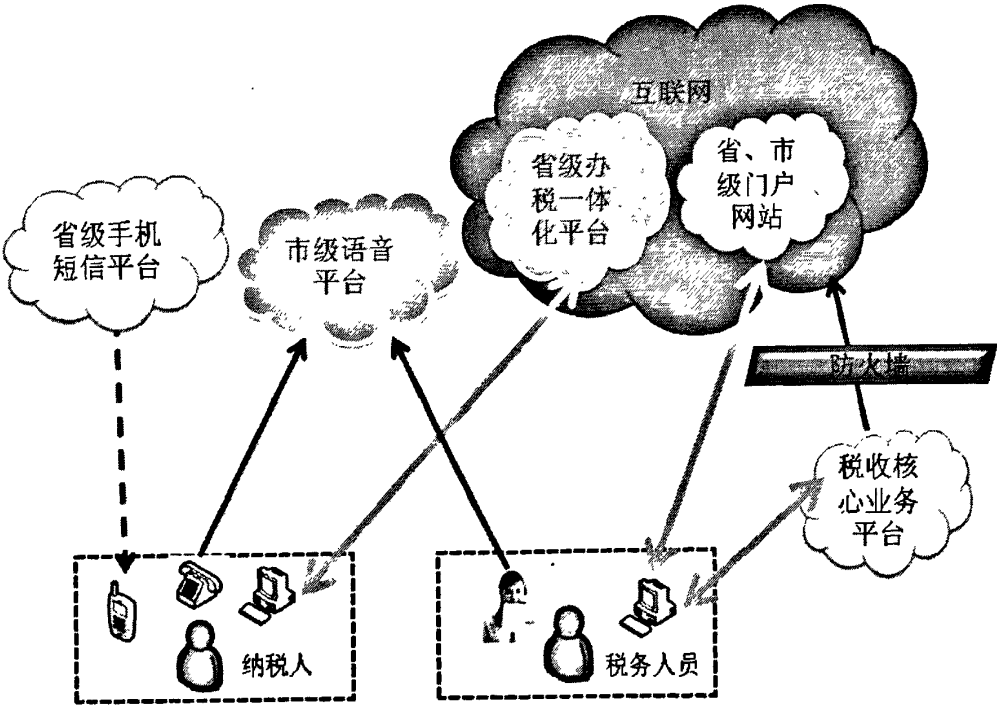


图 3-1 某省国税系统现有服务体系

Figure 3-1 the existing service system in a province State Tax System

### 3.1.1 纳税人面对的问题

1. 应用系统使用、维护有难度。国税部门虽然免费提供应用系统给纳税人使用,可是这些应用系统需要经过从网上下载、安装、升级、数据备份等处理,给纳税人带来便捷的业务处理的同时也带来技术处理方面的负担。纳税人需要频繁进出不同的应用系统、多次输入用户名和口令,需要负担聘请技术人员费用、承担应用系统甚至操作系统崩溃的痛苦和数据丢失的危险。在纳税人眼中,这些应用系统成了“鸡肋”,还有人甚至认为这是税务部门把数据采集的工作量强加给了纳税人。

2. 数据难以共享有“孤岛”。各应用系统的业务需求一般由各个部门在不同时期根据不同的业务流程和内容提出的,缺乏统一的分析、设计、开发标准,数据管理系统不一,数据共享比较困难,普遍存在“信息孤岛”现象。例如,由于征管业务系统与手机短信群发系统不能数据共享,税务部门的手机短信平台一般只发送事项通知而不能接收短信,更不能根据短信内容做出相应业务处理。业务系统生成的催办通知,需要系统管理员二次导入短信平台,否则群发时部分纳税人会收到与自己无关的催办业务的短信。

3. 数据量逐年增加导致响应速度下降。目前,国税系统各应用业务一般采取服务由省级平台提供、数据由省级集中存储,这样会导致数据量一般每年以数百 GB 甚至 TB 级的增长,纳税人在进行数据查询或写入操作时会感觉响应速度逐渐变慢。税务部门必须像银行部门一样定期进行已有数据迁移、异址存储来保证服务响应速度或者另外设置查询机将信息查询和数据写入进行人工分理,但是这样的结果必然导致硬件投资增加和维护另一套“克隆”系统的多余。

4. 应用系统大量前置导致服务手段单一。大量应用系统由纳税人使用,税收业务办理存在对计算机及网络的依赖,导致了税务部门服务手段比较单一,目前一般仅能通过互联网办理纳税申报、报税认证等业务,其他如税务登记、发票领购业务则只能到办税服务厅上门办理,手机网络、手机短信、语音服务、自助办税终端系统(简 ARM)机等尚未得到充分利用。在提高政府效能和服务质量要求下,税务部门提供的服务手段还需要不断拓展,向银行、电信等服务行业学习,促进服务手段的多元化。

### 3.1.2 税务部门的难题

在现有服务体系框架下,税务机关面临越来越多的问题。

1. 服务器日渐增多且资源利用率不高。国税部门的信息化建设发展很快,各省、各市自行或委托开发的应用系统非常多。但是在“金税工程”三期实施之前,由于缺乏系统的规划以及信息技术发展程度的制约,各应用系统业务逻辑不统一、编码不统一、部署平台不统一,数据无法共享且都具有重要性。因此,每部署一个新的应用系统,都配置独立的服务器和备用服务器,以保证数据安全和业务畅通。随着业务的不断发展及信息化建设的深入,应用系统的建设越来越多,机房里逐渐摆满了各个时期采购的不同的应用系统服务器。为了满足日益膨胀的系统服务器,各省、各市国税局信息中心不得不再购置机架甚至扩建机房,以满足信息化建设的需要。在服务器数量日益膨胀的同时,却发现大部分服务器的CPU、内存、硬盘、IO等资源平时利用率都不高,从未出现资源使用高峰时段或高峰时段持续较短。在资源闲置的时候,这些资源又无法分配给部分高利用率的服务器使用。

2. 运营成本逐渐加大。为了保证应用系统运营,国税部门投入大量资金购置各类服务器、服务器中间件和软件使用许可,每增加一台服务器,都需要花费数万甚至数十万元。同时,投入使用硬件设备增加又带来电力消耗的增加和机房温度的升高,维持机房温度的设备运行时间延长,日常运行成本不断加大。

3. 技术人员维护工作量大。以省国税局信息中心为例,信息中心机房运行有约20个应用系统,近60台各类服务器。系统管理员为了保障业务系统全天候正常运行,疲于维护,有时还需要新部署测试环境,工作量大。此外,随着办公自动化(OA)程度逐渐提高,办公楼各部门已经遍布了几百台PC机,为了保障各部门日常工作顺利开展,技术人员经常辗转于各个办公室安装应用系统、处理软件故障。

## 3.2 纳税服务云平台

基于云计算的纳税服务平台以跨平台的JAVA语言为开发工具,引进开源基础构件与开发应用构件相结合,以虚拟化技术整合现有硬件、软件资源,以分布式计算和分布式存储为基础,实现各应用系统与现有税收业务核心平台的数

据互联互通并逐步整合，建设支持网络、电话、手机、自助办税终端系统等多终端的综合服务平台软件系统，并以纳税人需求为导向逐步提供更加丰富的应用服务，不断满足纳税人多元化、便捷、高效的服务需求。基于云计算的纳税服务平台框架如图3-2所示<sup>[1]</sup>。

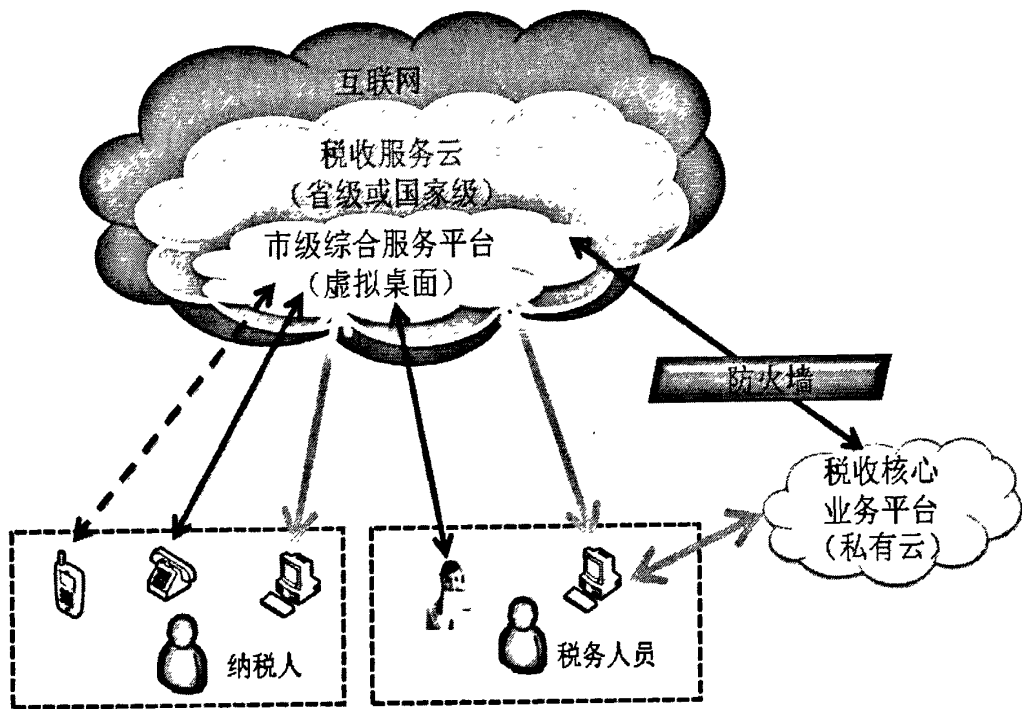


图 3-2 基于云计算的纳税服务平台框架

Figure 3-2 the tax service platform framework based on cloud computing

云计算自身的技术特点决定了它非常适合目前纳税服务信息化的发展现状，能够将现有服务体系实现质的飞跃。主要有以下优点：

1. 系统运营成本大大降低

应用云计算技术，可以将大量的、廉价的现有闲置资源作为节点加以有效利用，可以利用服务器虚拟化将硬件设备、应用软件在不同程度上进行整合，从而节约了大量的设备采购、软件使用费和电力等运行成本。

2. 更大程度上实现资源共享

通过云计算模式，各级税务机关将原来分散在各自单位的硬件、软件资源集中起来形成基础设施资源池，从而可以分享由大量节点连接在一起而形成的强大的计算能力和存储能力，而不必考虑硬件扩容、软件运维等压力。同时，

硬件资源的集中使应用系统的数据集中和共享成为可能和必要,可以有效避免信息孤岛的形成。

### 3. 降低前端运行维护要求

为了确保各类应用系统的正常平稳运行,国家、省、市、县级税务机关都要设置信息中心科室,配备大量技术人员,主要负责各类应用系统的推广、培训、运维和硬件检测、维修或报修等工作。在云计算模式下,服务器运维和各应用系统配置工作由云服务端统一管理,通常对用户终端的硬件配置要求不高,也不要求在客户端配置。因此,相关技术人员不必过多考虑硬件、软件升级、配置工作,降低了工作难度和强度。

### 4. 确保服务器的可靠运行

在现行运维模式中,应用系统通常会安装在某服务器上,一旦服务器出现故障,轻则不能提供正常服务,重则会导致数据丢失。在云计算模式中,数据多副本容错、计算节点同构可互换等技术提供了高可靠性保障。即使某台服务器出现故障,也可以将某台服务器中的数据快速克隆技术到别的服务器上提供服务,从而真正实现不间断服务。

## 3.3 云平台实现方案

由于税收数据可能涉及到国家经济运行状态和纳税人商业机密,因此,税收部门的云计算平台只能采取私有云的部署模式,避免资源、服务和数据由第三方托管而带来的潜在的危險或威胁。

纳税服务云平台的建设大致分两个阶段,第一阶段是以实现基础设施虚拟化或应用平台虚拟化为主,包括硬件设备虚拟化、桌面虚拟化等;第二阶段主要是以分布式计算、分布式存储为基础,在统一的标准下,不断部署新的税收业务应用系统,直接或间接依托第三方电信运营商提供给纳税人使用。在云计算所涉及的技术中,虚拟化是云计算的基础和前提,是云计算的第一步<sup>[12]</sup>。

### 3.3.1 实施进程

纳税服务云平台建设的第一个时期是非常重要的,必须保证应用服务连续和已有数据安全。一般认为从现有的信息系统过渡到私有云需要经过三个阶段。



#### 1. 统一计算平台

从现有信息系统向云计算系统过渡的第一步是硬件资源的集成，就是把现有的服务器、存储和网络设施等硬件以虚拟化技术集成在一起。虚拟化是让应用和物理硬件之间分离开的重要步骤，能够无缝在不同物理硬件间迁移，是虚拟化带来的直接好处，有了第一步的虚拟化，服务器、存储和网络都获得了极大的灵活性，这不仅能够进一步降低总体拥有成本，也是私有云能够平滑的构建“云”的必要条件。实际上，目前厂商提供的大多数私有云解决方案都处于这个层次，其优势在于能够节约用户安装、测试的时间，缩短业务上线周期。

#### 2. 管理集成

在虚拟化之后，建设方需要考虑“如何利用这样的虚拟化环境提供便捷、丰富的管理功能？”和实现端到端的整体自动化。要对管理进行整合和集成，并把符合安全、可信、合规的管理功能融入到虚拟化的基础架构中去——这在物理世界中是很难的，第一步的虚拟化工作就是为了完成这一步而做的准备。在这个阶段，已经有厂商提供一些管理平台软件，超越系统与系统之间管理的界限，不仅仅能够减少安装测试费用，还能够减少管理的复杂度和成本。

#### 3. 服务集成

服务集成阶段主要是从管理的集成过渡到服务的集成，在系统硬件架构上集成和提供一些用户的服务应用，使用户真正向“云”过渡。

服务集成阶段需要考虑的是如何进行“隔离”——这是类似多租户环境的概念，EMC公司高级副总裁兼首席技术官Jeff Nick认为这是建设私有云的最后一步，十分重要，同时也是虚拟环境中不同于物理环境的一步，在物理环境中，如果需要将不同的应用环境隔离，是一件非常复杂的工作，全面的防数据丢失的应用是很困难的，为了确保应用和操作系统的功能，管理员要将应用系统和堆栈从操作系统工程中隔离出来。在虚拟化环境中，操作系统和应用系统封装在一个虚拟机中，它们存在于一个共同的虚拟环境。

服务集成阶段的最终目标是各个应用部署和各个私有云的用户都能够在一个多租户的环境下共同利用一套物理资源。

### 3.3.2 风险点

云计算的技术基础其实就是网络计算、分布式计算、并行计算、分布存储等技术。这些技术的共同点都是为了满足海量计算或海量存储，而利用计算机集群这种横向扩展方法将数据分布或者计算任务分布到多台机器上，达到协作完成任务的目的。云计算的运行环境一般是廉价的普通PC 服务器，甚至是PC机，而且往往是异构机器——其存储部件多是SATA 硬盘，而非高级磁盘阵列。相对而言，传统的集群系统——如并行数据库等设计的运行环境很多是在专业服务器上完成，而且是同构的一批机器。

这种运行环境的区别决定了云计算架构设计的很多策略与传统分布系统、集群系统的架构设计有所不同。云计算环境首先需要考虑廉价机器的异常故障情况要远多于专有硬件平台，其次还需要考虑异构环境的速度不匹配问题也会对并行任务执行影响不小。

1. master单点故障。一般来说，云计算环境中没有特殊节点，都是对等的，这种无单点的集群结构，对故障容忍度更高，比如分散式杂凑表（Distributed Hash Table, DHT）架构的存储网。但是在目前实际运用中，Eucalyptus、hadoop、bigtable 等云服务属于master/salve 结构，系统中存在master 单点，因此部署时最好对master单点考虑高可用性（Hight Availablilty, HA），避免因单点故障引起的环境瘫痪。

2. 系统架构。云计算最大的优点是成本优势，设备便宜、通用，因此容易实现横向扩展。运行于云计算环境的集群最好能充分利用这种扩展成本优势，将系统设计成易于动态扩展的高可用性系统。如何设计易于扩展系统呢？目前最普遍的方式是share-nothing 架构。这种结构不同于分层架构，它需要将请求接受、处理和存储等都处于一个节点内完成（每个节点不一定是一个物理机器，可以只是一个虚拟运行容器），这样以来各个点之间不相互依赖，很容易进行线性扩容。并行数据库CouchDB、MongoDB等都属于这种架构。

3. 存储。云计算环境的存储由于没有专门的磁盘阵列（一般都支持RAID），而且节点系统可靠性不高，因此往往设计上需要使用软件方式实现多副本存储，以提高数据的安全性（当然也起到了负载均衡作用）。比如当前云环境中比较流

行的GFS 文件系统, dynamo 存储系统都是如此。

4. 计算中间结果的持续化。在传统的集群环境中, 节点故障率比较低, 因此各个节点之间常使用pipeline 等形式传送中间结果, 以保证更高的性能, 但是如果遇到处理节点故障则需要重新开始计算。云计算环境的数据计算(比如并行数据库)等, 需要尽可能的将任务划分, 把中间结果持续化, 避免某个处理节点故障后, 由于没有中间数据, 而不得不把整个计算任务重新开始。当前云计算中hadoop 的映射/规约(map/reduce)任务便是这种设计。

5. 异构环境引起的运行性能下滑。对于异构机器来说, 即使采用虚拟化技术, 由于调度方面的原因, 同一宿主机上的各同配置虚拟机执行性能也有一定的差别, 在云计算环境中各任务执行速度差异显著。因此, 往往需要跟踪发现执行缓慢的任务, 以便把任务调度到更快节点上重新做, 或者进行任务迁移等。目前Hadoop充分考虑了异构集群执行能力差异问题, 已经实现了这种特性。但是目前很多设计在同构机器上运行的并行数据库, 并不会重调度或任务迁移, 因此当到云计算这种异构环境运行时, 任务的最终执行会被最慢的那个拖累, 造成整个任务执行性能下滑。

### 3.3.3 工作任务

Ubuntu企业云(UEC)可以提供必要的工具帮助用户在自己的IT基础设施上创建私有云, 用户可以在安全环境下享受云计算带来的好处。UEC整合了一系列的开源项目, 包括KVM、Libvirt和Eucalyptus。Eucalyptus是Amazon EC2的一个开源实现, 与Amazon EC2的接口兼容。由UEC具有免费开源、容易安装配置、镜像和实例易于管理等特点, 本文以UEC为云计算环境研究应用系统移植需要完成的工作。

1. 硬件移植。特殊硬件需要软件化实现。UEC提供的Eucalyptus平台是使用KVM或XEN按需提供虚拟机, 用户可以在这些虚拟机上安装各种应用系统(预先制作好虚拟机镜像并上传注册), 但是并不允许在虚拟机环境中使用特殊硬件(如Teradata中联接各计算点的数据总线)。

2. 数据库移植。移植数据存储层到Eucalyptus的实例上。如果使用的是关系数据库(mysql、oracle9i)等, 就需要利用弹性块存储(Elastic Block Store,

EBS) 做后台存储体，将关系数据库服务程序安装到Eucalyptus实例中。同时，必须设计好数据冗余机制，防止某个实例发生故障时数据不丢，并且可以提供不间断服务。某些并行数据库的数据存储是利用磁盘阵列完成，而云环境并没有这种磁盘阵列。最直接的替代方法就是采用EBS服务——挂载外部存储磁盘到Eucalyptus系统上，这样可做软RAID 方案来代替磁盘阵列。如图3-3所示。

如果应用系统主要用于存储语音、图像、视频、Office文档等非文本信息时，可以采用开源的MongoDB、CouchDB、Cassandra等面向文档或大数据量对象处理的数据库实现分布式文件存储，能够很好地与云计算平台集成。

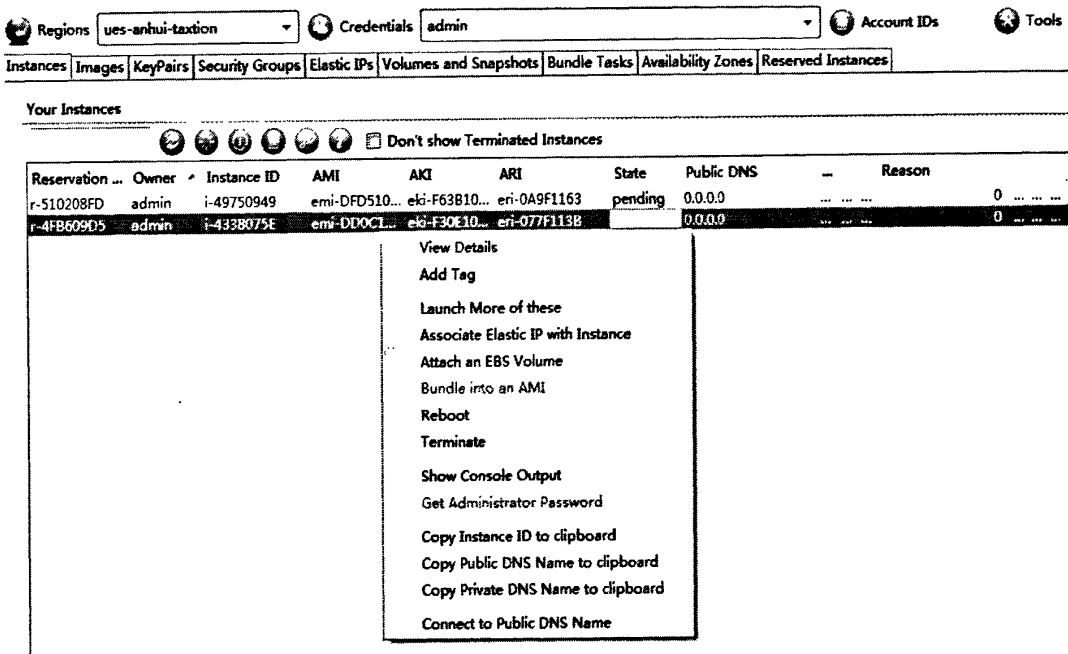


图 3-3 挂载EBS卷

Figure 3-3 attach an EBS volume

3. 应用服务移植。将应用程序和Web Server等部署到Eucalyptus的实例上（将在以后章节讨论），如果访问量大则可以将多个实例部署成集群。同时，也需要考虑到虚拟机实例可能会出现故障，设计故障转移策略减小风险，例如镜像克隆备份或跨地区部署物理服务器等。

## 第四章 纳税服务平台设计

纳税服务平台的功能从服务对象方面来说,首先能够提供纳税人办理税收业务和咨询、投诉等方面的功能,其次能够提供税务机关在业务管理方面的功能;从所提供的信息类型来说,则需要提供对文本、语音、图片、图像等信息的记录、管理和共享。由于税收政策的变动、纳税人需要的不断发展等原因,一劳永逸地设计好纳税服务平台的所有功能是不太现实的。

就现有服务内容和技术手段而言,基于云计算的纳税服务平台(以下简称:纳税服务云)应该是一个免填单、可定制、易部署、快响应、支持多终端的综合服务平台。

### 4.1 总体架构

基于云计算的纳税服务平台建设模式的重要特点就是依托国税自己建设的广域网络、存储资源、计算资源部署应用系统为纳税人和税务机关提供业务办理、存储服务,基于SOA的应用系统一般由国税部门自行开发或以软件外包形式委托服务商开发。

经过分析论证,结合云计算技术和国税税收业务的特点,本文提出了纳税服务云计算总体架构,如图 4-1 所示。

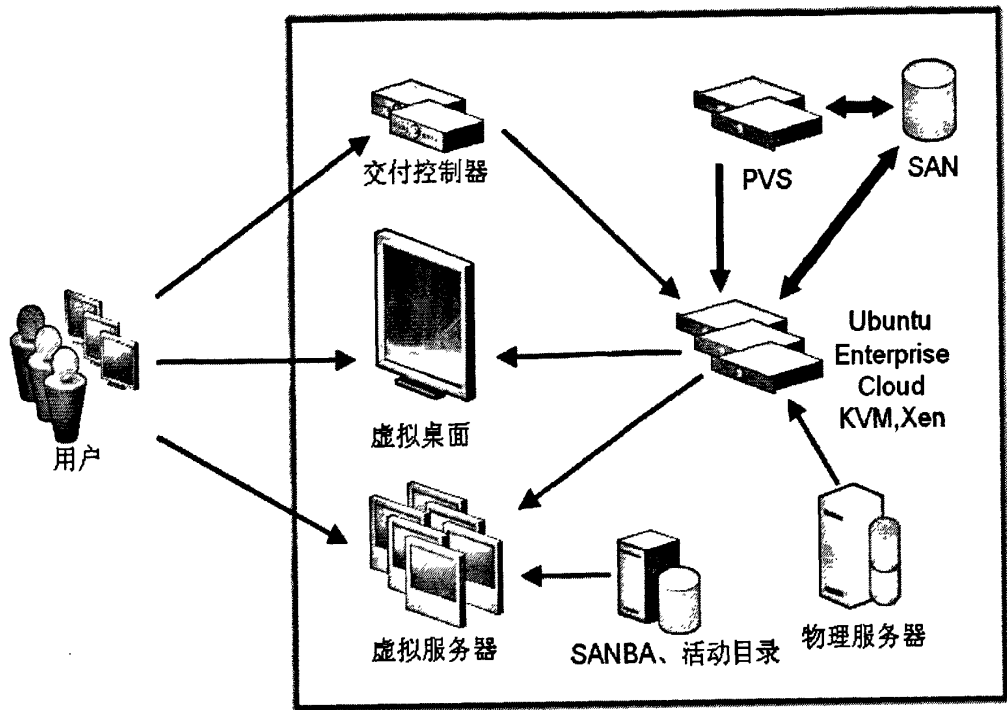


图 4-1 纳税服务云计算平台的总体架构

Figure 4-1 the overall structure of Tax Services cloud computing platform

## 4.2 需求分析

纳税服务平台的功能从其所服务的对象来说，首先要能为纳税人办理税收业务、查询税收法规、税款自动划转、投诉建议、语音、短信等服务，其次要能为税务部门提供日常管理、信息共享等服务。存储的数据类型一般有数据库、文本、图像、视频、音频等。从国家税务总局下发的《全国税务系统 2010-2012 纳税服务规划》来看，纳税服务云平台建设涉及纳税人、税务部门、通信服务商、银行、国库等多个行业，整体设计和开发工作很难一蹴而就，而是需要一个长期的过程。目前纳税人常用的或迫切需要税务部门提供的功能主要涉及实现多元化申报缴税和税收业务档案电子化存储方面。

根据当前纳税服务工作的要求和发展趋势，基于云计算的纳税服务平台主要包括网上办税、手机短信、电话语音、自助办税终端、税收征管档案等五个交互式服务子平台，并可在将来根据需要扩展。

### 1. 网上办税服务

建设省级集中的网上办税平台，形成以省局网站为主站，各市局网站作为主站频道（虚拟目录）的网站群。该平台能够实现网上认证、远程报税、网上申报和自动缴纳税款、预订（购买）发票、政策查询等交互式涉税业务。同时，能够为不同纳税人提供不同的虚拟化桌面，消除纳税人安装、维护各类应用系统客户端的烦忧。

### 2. 手机短信服务

现有短信服务平台以市级为单位，独立于其他应用系统之外，并且无法对接收短信进行处理。新的短信服务平台应该像现有税收业务核心系统一样实行省级业务、数据集中，实现与业务核心系统数据共享，向电信运营商开放系统接口，实现纳税人电话申报纳税，实现短信查询、办理业务，并且根据业务核心系统产生的待办信息自动通知消息对象。

### 3. 电话语音服务

各地税务机关现有的电话语音服务平台以市级为单位，大部分是以人工接入为主。新的服务子系统将采用计算机电话集成（CTI, Computer Telephony Integration）技术和桌面虚拟化技术，由人工接入加单个PC的运营模式转变为自动队列加虚拟机模式。各市级呼叫中心的终端桌面配置、应用系统管理、数据存储等主要工作省级数据中心的云服务器完成，从而大大减少市级呼叫中心的部署、响应时间。应用系统提供的功能主要是税收法规咨询、意见建议反馈等。

### 4. 自助办税终端（ARM）服务

自助办税终端作为一种自助服务的平台，能够在满足税收管理需求的前提下，为纳税人提供安全便捷的办税服务。它借鉴银行、电信等行业的终端处理方式，可以放置在办税服务厅的公共服务区，为纳税人提供 7\*24 小时服务。自助办税终端能够支持涉税信息公开，便于行政相对人的查询，最主要的是要实现与业务核心系统的数据共享，实现纳税申报、发票发售、发票认证、报税、缴纳税款、法规查询等服务。

### 5. 税收征管档案服务

税收征管档案平台应该满足存储纳税人提供的各类证件资料、征纳过程中产生的各类文书资料，并提供各类查询服务，在税收业务办理过程中实现一次

采集、永久存储、资料共享。

根据以上需求分析,纳税服务云平台的用户分为纳税人、税务人员、系统管理员和非注册人员(Guest)。已注册用户在相应的角色范围内可以使用交互式操作,非注册用户只能浏览公开的资料信息。

### 4.3 存储设计

在网上办税、手机短信、自助办税终端等子平台中,应用系统一般存储的是文本信息,适合选用关系型数据库,例如Oracle,以便使用结构化查询语言在存取属性较多的数据记录时获取更高的效率。

在电话语音和税收征管档案子平台中,存储的数据主要是语音、图像、视频、Office文档等非文本信息,应该选用非关系型数据库,例如基于分布式文件存储的数据库MongoDB。

MongoDB是一个介于关系数据库和非关系数据库之间的产品,存储数据非常方便,具有高性能、易部署、易使用的特点,目前是非关系型数据库中功能最丰富,最像关系数据库的。MongoDB旨在为WEB应用提供可扩展的高性能数据存储解决方案,支持的数据结构非常松散,采用的存储形式为BSON(Binary Serialized document Format),可以存储比较复杂的数据类型。Mongo最大的特点是其支持的查询语言非常强大,其语法类似于面向对象的查询语言,几乎可以实现类似关系数据库单表查询的绝大部分功能,而且还支持对数据建立索引。

### 4.4 网络选择

根据税收业务特点和不同规模纳税人对数据安全的需求,数据传输方式一般有虚拟专用网络(Virtual Private Network, VPN)、互联网、无线应用协议(Wireless Application Protocol, WAP)三种。

1. 虚拟专用网络(VPN)。虚拟专用网络一般适用于数据传输量较大、对数据安全要求高的大型企业。虚拟专用网指的是在公用网络上建立专用网络的技术。整个VPN网络的任意两个节点之间的连接是架构在公用网络服务商所提供的网络平台(如Internet、帧中继、ATM等)之上的逻辑网络,并没有传统专



用网络需要的端到端的物理链路,用户数据在逻辑链路中传输。使用虚拟专用网络可降低成本、保证传输数据安全可靠、连接方便灵活、完全控制自己网络安全。

2. 公共互联网。这是日常生活中最常见的数据传输网络,成本低廉,安全性较差,数据传输速度得不到保证,但具有易于部署、使用的特点。公共互联网一般适用于数据传输量较小的中小型企业,通过税务机关发放的 CA 认证保证一定的安全性。

3. 无线应用协议。无线应用协议(WAP)于1998年初公布,是全球性的开放标准。WAP的目标就是将 Internet 中先进的应用业务和丰富的信息及引入到无线终端之中。它根据无线网络的特点进行性能优化设计,把 Internet 中有关的规范、协议引入到无线网络。WAP 仅仅需要 WAP 代理服务器和无线终端产品的支持即可,无需对已有的移动通信网络协议做任何修改。WAP 尽可能地减少对无线终端设备的 CPU、RAM 和 ROM 等资源的占用,它通过提高网络的功能弥补无线终端设备的缺陷。基于 Internet 标准的微浏览器技术可以使无线终端设备根据自己设定的输入、输出方式来决定显示信息的方式。这样,从最简单的手机、功能相对丰富的 PDA 以及笔记本电脑等各种终端都能接收到相同的信息和资源。无线应用协议一般适用于小规模纳税人和个体工商户,可以使用各种终端办理简单的税收业务。

### 4.5 功能实现

由于纳税服务平台的复杂性和较大的工作量,暂以税收征管档案管理子系统为例进行系统开发、测试、部署。税收征管档案中的静态资料(如工商、税务登记,银行账户等)一次性采集、扫描、存储,多次调阅使用,是减轻纳税人负担、提高服务效率的有效措施,也是降低税务机关档案存储、调阅成本的必由之路。资料存储格式多为图片、视频、语音、文档等二进制格式。

#### 4.5.1 开发环境

系统开发工具为 Weblogic Workshop,开发语言为 Java,前端表示层采用 Beehive+JSP,应用服务器中间件采用 Weblogic Server,数据存储层采用 MongoDB。

BEA Weblogic Server 和 Workshop 是由著名的 Java 中间件软件公司 BEA Systems (2008 年 1 月 16 日被 Oracle 并购) 推出, 是 BEA SOA 的重要基石。BEA Workshop 10 支持 J2EE5, 能够有助于加快并从根本上简化应用开发的过程。它提供一种集成的对象/关系映射 Studio 以及面向 Eclipse 的 BEA Kodo、EJB3/JPA、OpenJPA 和 Hibernate 等工具。Workshop 10 还提供了独特的 AppXRay 技术和所见即所得(WYSIWYG)的开发功能, 它们为众多开源和行业标准框架而设计, 包括 Java Server Faces (JSF)、Struts/Tiles、Spring、JSP/JSTL、EJB3/Hibernate、Apache Beehive 及其他框架, 有助于进一步简化开发过程、提高整体生产力。BEA Weblogic Server 是一个功能丰富、基于标准的应用服务器, 为构建可靠、可伸缩和可管理的应用程序提供了坚实的基础, 为开发和部署业务驱动的应用程序提供了必需的底层核心。

#### 4.5.2 关键功能

在系统中涉及的数据分为结构化数据和非结构化数据, 结构化数据在 MongoDB 中采用 BasicDBObject 类组成的数据集存储, 读写和查询操作易于处理。系统还要经常涉及到各类文件上传、下载操作, 文档资源数量众多、类型不同、大小不一, 写入、查询、读取效率对用户体验效果的影响最大, 如果解决不好将会是整个应用系统的瓶颈。因此, 在系统实现过程中, 资源上传、查询、下载功能将作为关键部分来处理。

##### 1. 上传资源

税务部门的操作人员可以上传文档、图像、视频等资源, 资源上传页面如图 4-2 所示。操作员在“纳税人代码”文本框内输入部分代码, 在模糊查询到的纳税人信息中选择, 然后选择好正确的资料类别、所属时间等信息, 将资料平铺在拍摄设备之下(如摄像头), 点击“存储”按钮即可完成图像拍摄、上传操作。

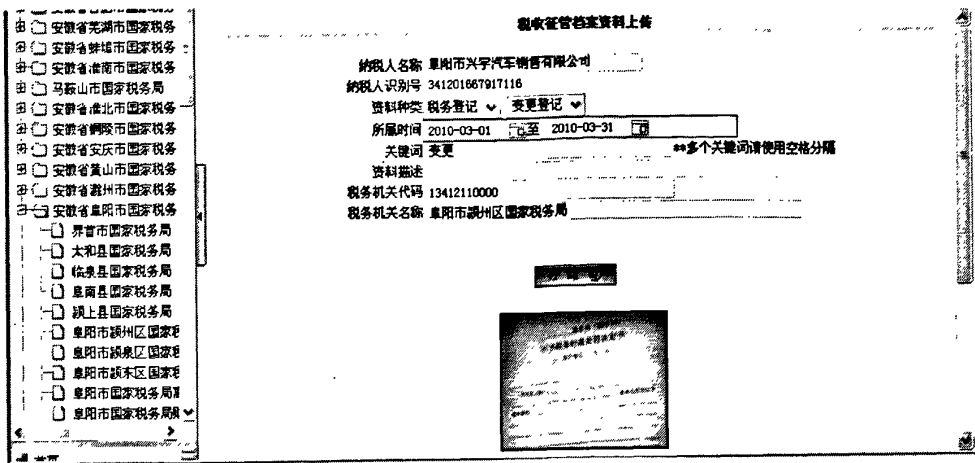


图 4-2 上传资源页面

Figure 4-2 Upload resources snapshot

上传功能的主要通过 Beehive 框架的 fileUpload 标签、FormData 控件、MongoDB 的 Java API 实现。资源的类别由 FormData 根据上传文件的扩展名自动识别，并将 MIME (Multipurpose Internet Mail Extensions) 类型保存在数据库中。写入数据所用到的类主要是 InputStream、DB、GridFS、GridFSFile 等。如果上传的文件达到或超过数据块 (chunks) 的最大容量时，会被切分到 2 或多个 chunks 中。当一个片 (shard) 包括数据过量时，chunks 将会被迁移到其他的 shard 上

资源上传功能的主要代码如下：

```
public void saveDocument(InputStream in, DocumentBean document)
{
    GridFS f = getGridFS();
    try {
        GridFSFile mongofile=f.createFile(in, filename);
        mongofile.put("filename", document.getFilename());
        mongofile.put("contentType", document.getContenttype());
        mongofile.put("uploadDate", document.getDate());
        mongofile.put("nsrsbh", document.getNsrsbh());
        mongofile.put("nsrmc", document.getNsrmc());
        mongofile.put("czry_dm", document.getCzry_dm());
        mongofile.put("zlzl_dm", document.getZlzl_dm());
        mongofile.put("ssrq_q", document.getSsrq_q());
        mongofile.put("ssrq_z", document.getSsrq_z());
        mongofile.put("gjc", document.getGjc());
        mongofile.put("zlms", document.getZlms());
```

```
mongofile.save();
} catch (Exception e) {
    e.printStackTrace();
}
}
```

2. 查询资源

查询功能主要是根据设置的条件实现精确查询、模糊查询和嵌套查询（在查询的结果中进一步筛选）。资源查询页面如图 4-3 所示。

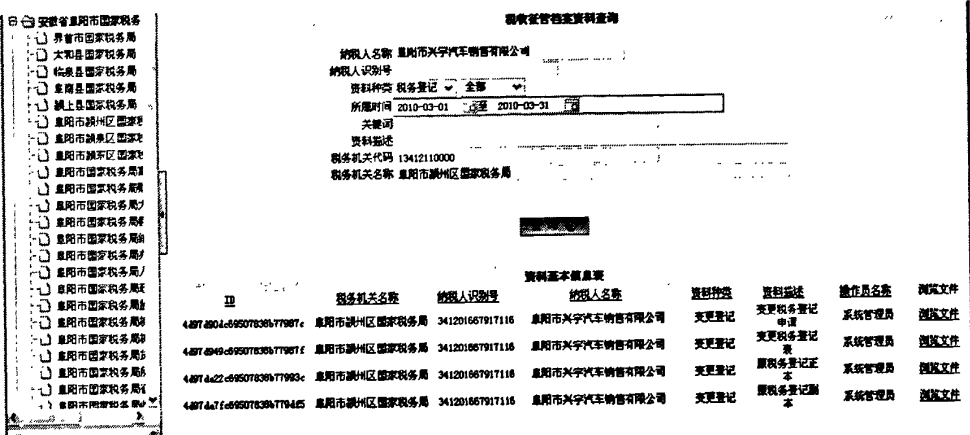


图 4-3 查询资源页面

Figure 4-3 Inquires resources snapshot

查询数据所用类主要是 BasicDBObject、HashMap、DB、GridFS、GridFSDBFile 等。系统根据设置的条件将查询到的结果显示在页面下方的表格中。

资源查询功能的主要代码如下：

```
//根据给定条件查询记录
public List<GridFSDBFile> findFiles(Map<String, Object> map) {
    BasicDBObject query = new BasicDBObject();
    query.putAll(map);
    List<GridFSDBFile> list = new ArrayList<GridFSDBFile>();
    list = getGridFS().find(query);
    return list;
}

//例如根据纳税人识别号查询记录
List<GridFSDBFile> list3 = new ArrayList<GridFSDBFile>();
GridFSDBFile gfsfile = new GridFSDBFile();
Map<String, Object> map = new HashMap<String, Object>();
```

```

map.put("nsrsbh", "341201667917116");
list3 = mongoDAO.findFiles(map);
if(list3!=null&&list3.size()>0)
{
    for (int i=0;i<list3.size();i++)
    {
        gfsfile = list3.get(i);
        System.out.println(gfsfile.getId().toString()+gfsfile.get(
"gjc"));
    }
}

```

### 3. 下载资源

用户首先根据需要设置条件查询出符合条件的资源,然后通过“浏览文件”打开新的窗口读取数据流到客户端。在 WEB.XML 中写入多个形如“<mime-mapping><extension>doc</extension> <mime-type> application / msword</mime-type></mime-mapping>”的语句块配置 DOC、JPG、MPEG 多种文件的 MIME(Multipurpose Internet Mail Extensions)类型,当数据流被传送到客户端浏览器时,浏览器可以启动相应的应用程序来处理这个输出文档。

资源下载功能的主要代码如下:

```

protected void downloadDocument(String id)
{
    try{
        DBObject dbobj = new BasicDBObject();
        Map<String, Object> map = new HashMap<String, Object>();
        map.put("id", id);
        dbobj.putAll(map);
        GridFSDBFile file = getgridFS().findOne(dbobj); //检索第一个
        Long filesize = file.getLength();
        getResponse().reset();
        //除掉原来默认的 text/html
        getResponse().setContentType(file.getContentType());
        getResponse().setHeader("Pragma", "No-cache");
    }
}

```

```
getResponse().setHeader("Cache-Control","no-cache");

getResponse().setDateHeader("Expires", 0);

getResponse().setContentLength(filesize);

ServletOutputStream servletoutputstream =
getResponse().getOutputStream();

BufferedOutputStream outbuff = new
BufferedOutputStream(servletoutputstream, 1024);

file.writeTo(outbuff);

outbuff.write(abyte0, 0, abyte0.length);

outbuff.flush();

outbuff.close();

} catch (Exception e) { e.printStackTrace();}

}
```

## 第五章 纳税服务平台部署

Ubuntu作为2004年底诞生的Linux操作系统，短短5年间发展迅速，目前全球已有1200万人在使用Ubuntu系统。2010年4月29日，Ubuntu 10.04 LTS

(Long-Term Support) 正式发布。Ubuntu 10.04进一步加强了Ubuntu的云计算特性，在 Ubuntu Server Edition中提供了Ubuntu Enterprise Cloud。

Ubuntu Enterprise Cloud (UEC)，即Ubuntu的企业云解决方案，是一个由Ubuntu提出的新的开源计划，旨在进一步简化基于 Eucalyptus 的云基础设施的部署、配置和使用。Ubuntu云计算是一个基础设施平台的概念，支持Amazon的EC2的标准。有了这个标准以后，用户既可以构建能够在Amazon的 EC2 基础设施上运行 Eucalyptus 的公共云，也可以构建能够在位于防火墙之后的数据中心内部的基础设施之上运行的私有云。这是到目前为止最为简单的安装和尝试使用 Eucalyptus 的方式。UEC 也是第一个允许在本地环境轻松创建云服务并进而利用云计算强大功能的开源项目。

UEC允许用户在私有的硬件设备上、防火墙以内构建像Amazon这样的云计算解决方案，可以很好地解决安全问题。另外，很多用户对计算的要求非常有弹性，时大时小，变化很大，UEC能够很好地满足这些根据需求去变换计算能力的用户要求。与微软和VMware相比，Ubuntu的云计算解决方案是完全开源的。

### 5.1 部署私有云

部署一个最小化运行的Ubuntu云计算基础架构 (cloud infrastructure)，至少要两个专用系统 (dedicated systems)，一个用于前端 (front end) 应用，包括云控制器Cloud Controller (CLC)、群集控制器Cluster Controller (CC)、Walrus Controller (类似亚马逊S3的存储服务) 和存储控制器Storage Controller (SC)；其他是一个或多个可以托管虚拟机的节点，包括节点控制器node controller (NC)。前端和节点对硬件需求参见文献<sup>[13]</sup>。

虽然Eucalyptus提供的镜像包括KVM、Xen和VMware内核 (Eucalyptus2.0支持)，但是Ubuntu企业云只使用开源的基于内核的虚拟机平台KVM托管虚拟机。

UES云控制器是一个前端服务应用程序，连接到它进行云服务的管理，它支持调用EC2 API，并提供了Web管理界面。群集控制器管理各个群集节点（VM）资源，并通过开源的libvirt库与每个节点/VM通信。Walrus提供类似亚马逊S3的存储服务功能。存储控制器与亚马逊的弹性块存储（EBS）类似，可以将连续的虚拟磁盘附加给正在运行的实例，激活资源存储虚拟化控制，为便于管理，每个集群都有一个存储控制器。

节点需要运行基于内核的KVM托管虚拟机，所以节点服务器的CPU必须支持虚拟化技术，并且已在BIOS中将“Virtual Technology”项设为“Enabled”，否则在安装过程中会出现错误（如图5-1），使用时无法启动虚拟机实例。

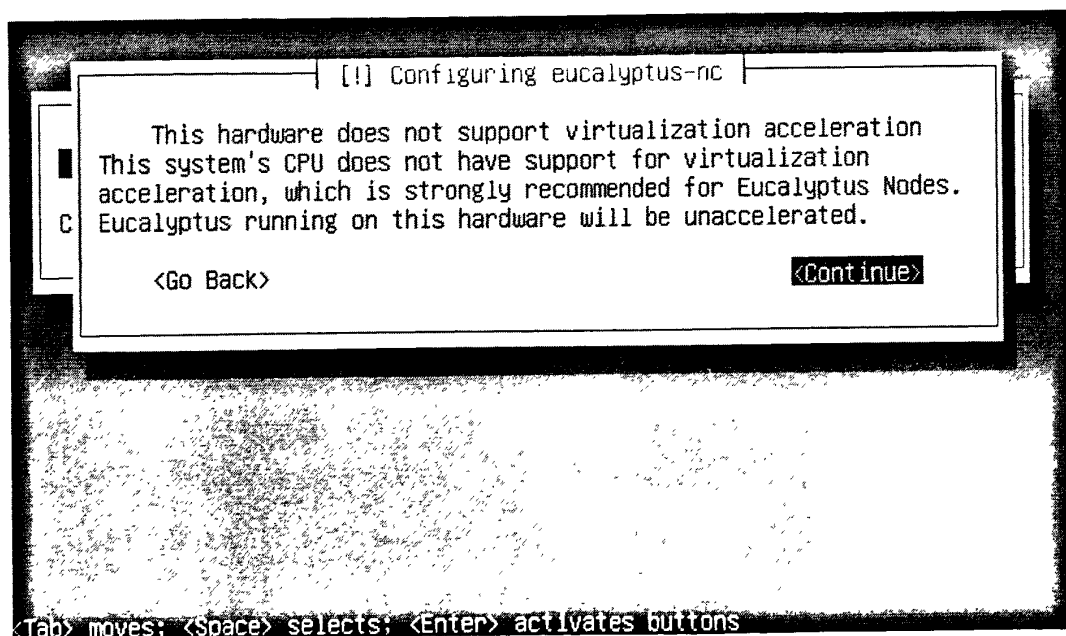


图5-1 CPU不支持虚拟化的提示信息

Figure 5-1 CPU does not support virtualization reminder

如果节点控制器和云控制器在同一局域网络上，节点会自动发现云控制器和群集控制器，安装成功后会自动完成节点注册。如果节点安装结束后，目录“/etc/eucalyptus/”不存在，则需要下载安装并配置eucalyptus-nc包。如果使用“euca-describe-availability-zones verbose”命令查看的信息中“availability zones”的“free/max”列数值没有增加，则说明节点无法被访问或者不支持虚拟化。

前端配置主要是根据硬件设施和提供服务需要而设置Cloud Controller



(CLC)、Cluster Controller (CC)、Walrus Controller和VM Types等项目的参数。如果前端安装结束后又改变了前端控制器的IP地址，一定要在“Configuration”中重新正确配置，否则在上传镜像 (Image) 时会出现结束处为“socket.error: [Errno 113] No route to host”的错误信息。

5.2 管理镜像

5.2.1. 下载镜像

在WEB管理界面中选择“Store”选项页，查找所需的镜像，下载并自动安装，如果安装成功相关信息会出现在“Image”选项页的列表中（如图5-2）。

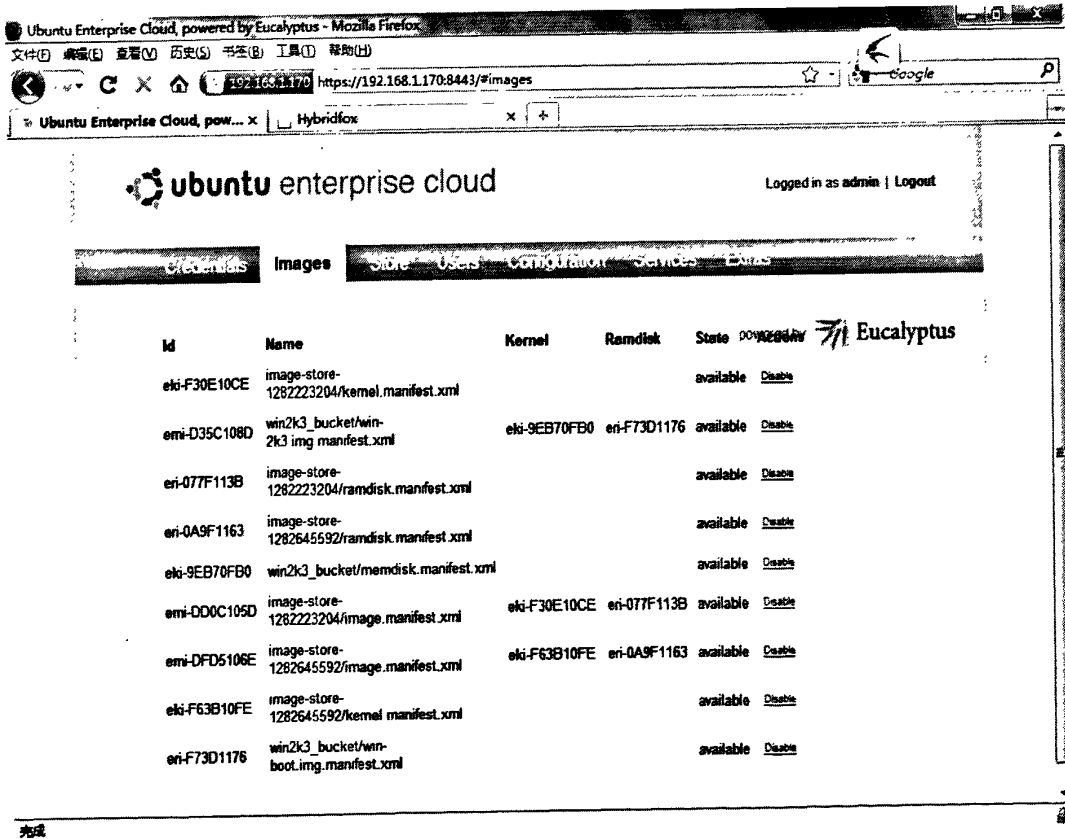


图5-2 上传注册的镜像列表

Figure 5-2 the list of registered images

5.2.2 自建镜像

实例终止运行后，对实例所做的安装应用系统、数据库等工作也将丢失。为了让实例能够达到指定的运行级别，必须到Eucalyptus网站下载合适的镜像

进行创建或修改（包括安装应用程序、配置映像、编写启动脚本以及配置服务等），然后使用euca2ools命令行上传到云控制器并注册为自己的AMI，从而使这些映像可以部署到云基础架构上。

由于Ubuntu所使用的Eucalyptus版本为1.6.3,尚不支持VMware虚拟化，但2010年8月推出的2.0版本不仅支持VMware，也支持Windows客户机。下表5-1是以windows系统为例自建镜像包的操作步骤：

步骤		命令行
创建磁盘镜像	创建新的镜像	<pre>\$ kvm-img create -f qcow2 win-2k3.img 20G</pre> <p>（Qcow2 是可以动态扩展容量的镜像格式）</p>
	安装OS	<pre>\$sudo kvm -m 1024 -cdrom /mnt/hda/win2003_cd1.iso -drive file=win-2k3.img,if=scsi,boot=on -nographic -vnc :0</pre> <p>如果提示：pci_add_option_rom: failed to find romfile "pxe-rtl8139.bin", 请下载 " pxe-rtl8139.bin" 到" usr/share/kvm" 目录下。</p> <p>如果有第二张安装盘：</p> <pre>\$ sudo kvm -m 1024 -boot c -cdrom /mnt/hda/win2003_cd2.iso -drive file=win-2k3.img, if=scsi, boot=on -nographic -vnc :0</pre>
创建boot disk	复制启动文件	<p>重启VM，并复制ntldr, ntdetect.com和sym_hi.sys 文件到 C:\bootfiles:</p> <pre>\$ sudo kvm -m 1024 -boot c -cdrom Win2003_cd1.iso -drive file=win-2k3.img,if=scsi,boot=on -nographic -vnc :0</pre> <p>输入 windows 的 Boot.ini 文件所需内容保存到 C:\bootfiles目录中的 Boot.ini文件。</p>

	创建启动盘	<p>1. 在CC控制器上, 执行以下命令行创建启动盘:</p> <pre>\$ dd bs=512 count=2880 if=/dev/zero of=win-boot.img</pre> <p>2. 以此盘作为软盘并启动VM:</p> <pre>\$ sudo kvm -m 1024 -boot c -drive file=win-2k3.img, if=scsi, boot=on, index=1 -fda win-boot.img -nographic -vnc :0</pre> <p>3. 在VM中格式化并复制C:\bootfiles目录中的所有文件到此软盘。关闭VM, 使用以下命令测试此启动盘:</p> <pre>\$ sudo kvm -m 1024 -boot a -drive file=win-2k3.img, if=scsi, boot=on, index=1 -fda win-boot.img -nographic -vnc :0</pre>
	网络配置	<p>1. 从Intel站点下载e1000网卡驱动, 然后重启VM, 并安装驱动。</p> <pre>\$ sudo kvm -m 1024 -boot a -drive file=win-2k3.img, if=scsi, boot=on, index=1 -fda win-boot.img -nographic -vnc :0 -net nic, model=e1000</pre> <p>2. 为了远程访问VM桌面, 需要在VM中打开“控制面板→系统→远程”并为所有用户开启“远程桌面”选项。</p>
创建 Memdisk		<p>如果在宿主机的目录 /usr/lib/syslinux 中没有 memdisk, 需要下载并安装:</p> <pre>\$ sudo apt apt-get install syslinux, 然后复制memdisk到windows文件所在位置, 即win-2k3.img和win-boot.img。</pre>
上传并注册镜像	注册 memdisk	<pre>\$ euca-bundle-image -i memdisk --kernel true \$ euca-upload-bundle -b win2k3_bucket -m /tmp/memdisk.manifest.xml \$ euca-register win2k3_bucket/memdisk.manifest.xml</pre> <p>(记下最后的命令的输出信息, 类似“eki-XXXXXXX”, 备用。)</p>

	注册 boot disk	<pre>\$ euca-bundle-image -i win-boot.img --ramdisk true \$ euca-upload-bundle -b win2k3_bucket -m /tmp/ win-boot.img.manifest.xml \$ euca-register win2k3_bucket/win-boot.img.manifest.xml</pre> <p>(记下最后的命令的输出信息, 类似“eri-XXXXXXX”, 备用。)</p>
	注册 磁 盘 镜像	<pre>\$ euca-bundle-image -i win-2k3.img --kernel eki-XXXXXXX --ramdisk eri-XXXXXXX \$ euca-upload-bundle -b win2k3_bucket -m /tmp/ win-2k3.img.manifest.xml \$ euca-register win2k3_bucket/win-2k3.img.manifest.xml</pre> <p>(将eki-XXXXXXX和eri-XXXXXXX替换成前面记下的真实值。)</p>
查看已上传镜像		<pre>\$ euca-describe-images</pre> <p>(记下输出信息中类似“emi-XXXXXXX”的数值, 备用。)</p>
安全组设置		<p>验证实例使用的安全组是否允许使用RDP:</p> <pre>\$ euca-describe-groups</pre> <p>如果未返回相关信息, 使用下面命令开启RDP:</p> <pre>\$ euca-authorize default -P tcp -p 22 -s 0.0.0.0/0</pre>
运行实例		<pre>\$ euca-run-instances emi-XXXXXXX -k mykey -t cl. xlarge</pre> <p>(将emi-XXXXXXX替换成前面记下的真实值。)</p>
访问实例		<p>从远程机器上, 执行“开始→程序→附件→远程桌面连接”, 输入该实例的IP地址即可访问。</p>

表5-1 Windows系统镜像打包操作步骤简表  
Table 5-1 Windows system image package steps Summary

5.3 管理实例

Ubuntu企业云使用Eucalyptus的euca系列命令来管理镜像实例。另外, 从1.6版本开始, UES提供基于Amazon’ s ElasticFox的Firefox插件HybridFox<sup>[14]</sup>

为使用、访问和监控云资源提供了一个简单的GUI，Ubuntu企业云Web界面和HybridFox相结合可以简化私有云的管理，而ElasticFox的最新版本与Eucalyptus不再兼容无法工作。

5.3.1 euca系例命令

Eucalyptus提供的euca系例命令包括euca-describe-availability-zones verbose（察看可用的资源）、euca-describe-instances（察看目前正在运行的虚拟机）、euca-bundle-image（创建镜像）、euca-upload-bundle（上传镜像到Eucalyptus）、euca-register（注册已经上传的镜像）、euca-describe-images（查看可用的镜像）等等，详情可参考文献<sup>[15]</sup>。

5.3.2 HybridFox应用

截止目前，尚无资料详细描述HybridFox的使用。使用HybridFox，可以非常方便地管理所有镜像，实例，安全密钥对，安全组，弹性IP地址范围，EBS卷和集群等，可以从一个镜像创建一个新的实例，通过ssh连接到这些实例，添加或编辑我们的密钥对，安全组，创建/删除或附加EBS卷。遗憾的是，HybridFox还不能从运行中的实例创建新的包，必须从命令行来完成这项工作。具体使用方法如下：

1. 添加“Regions”和“Credentials”（如图5-3），

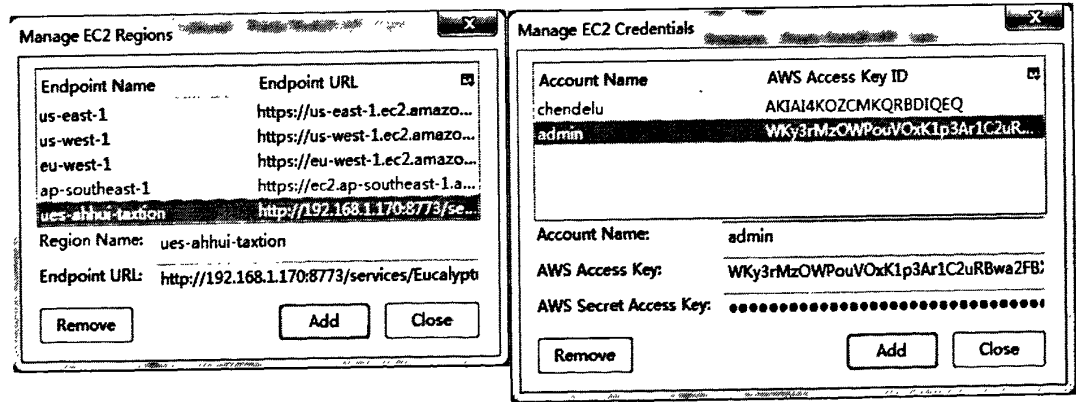


图5-3 管理Regions和Credentials

Figure 5-3 Management Regions & Credentials

2. 选择“Security Groups->Group Permissions”（如图 5-4），或者在云控制器（CLC）上使用 ec2-authorize 命令行工具（例如“euca-authorize default -P tcp -p 22 -s 0.0.0.0/0”）授予用户对 HTTP、FTP 等应用的端口访问权，用户才能通过这些端口访问各类应用。

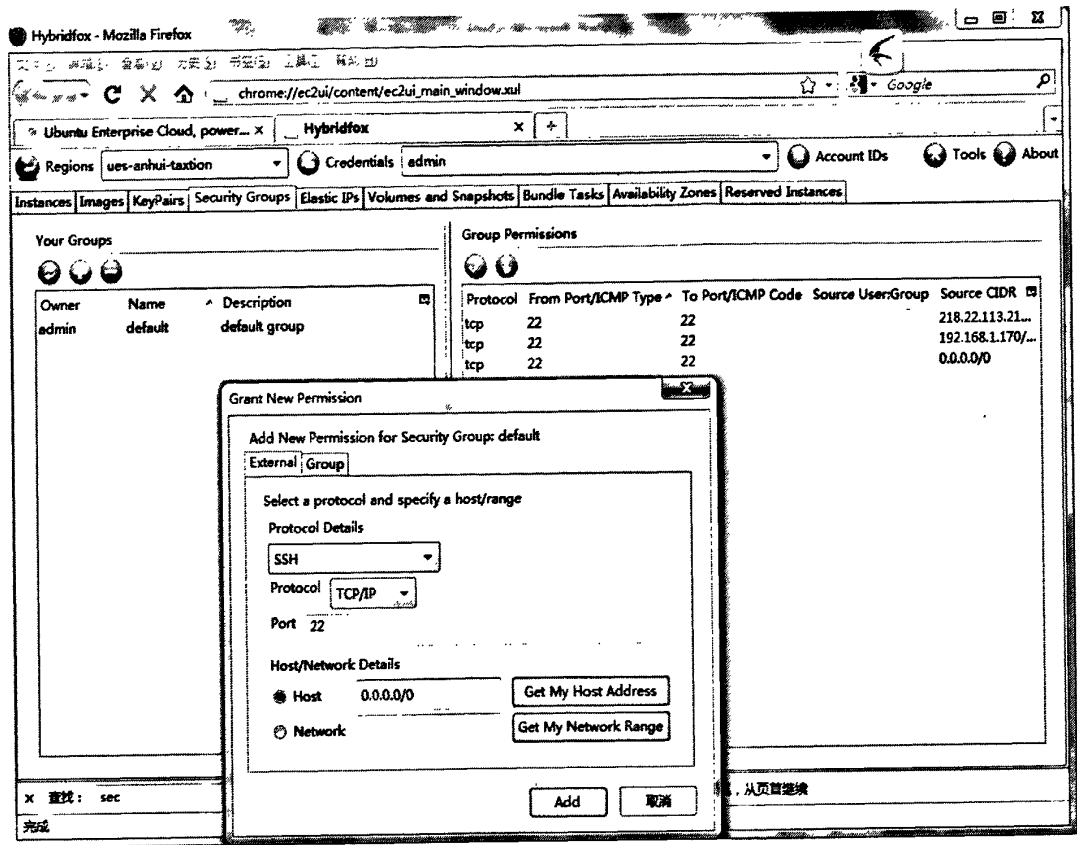


图5-4 工作组权限管理

Figure 5-4 Group Permissions

3. 为了能够运行、访问虚拟机实例 (Instances)，需要使用“KeyPairs”功能生成某用户的 PEM 类型的私钥（如图 5-5）。

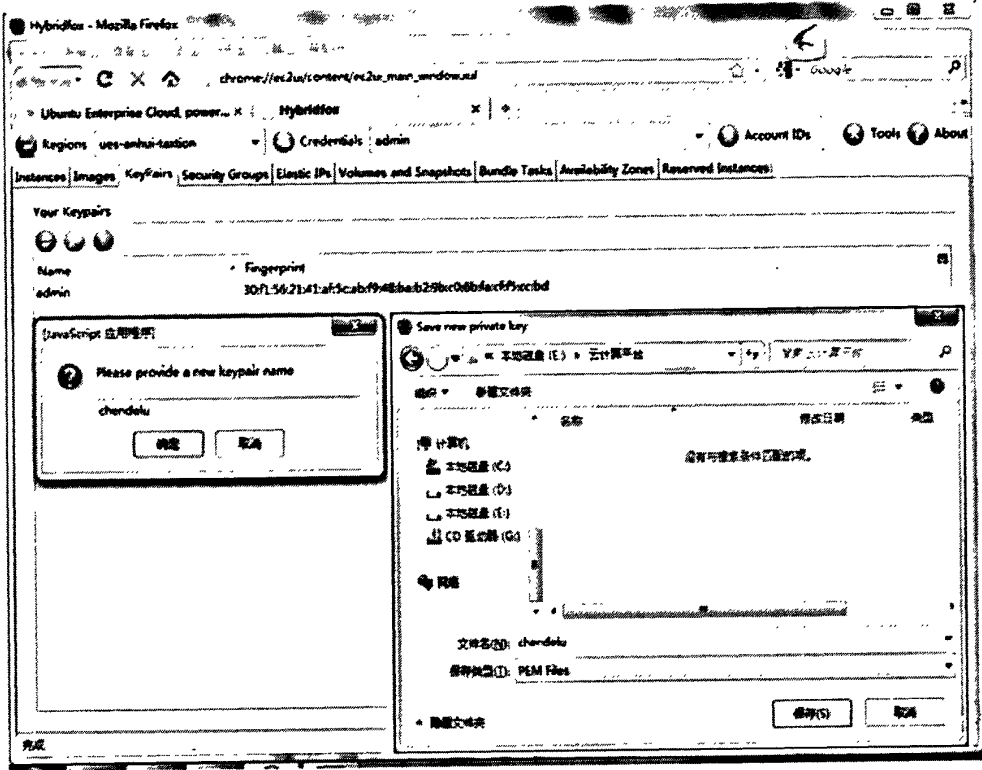


图 5-5 生成私钥

Figure 5-5 Private key generation

4. 选择 “Images->Launch instance”，运行虚拟机实例（如图 5-6）。

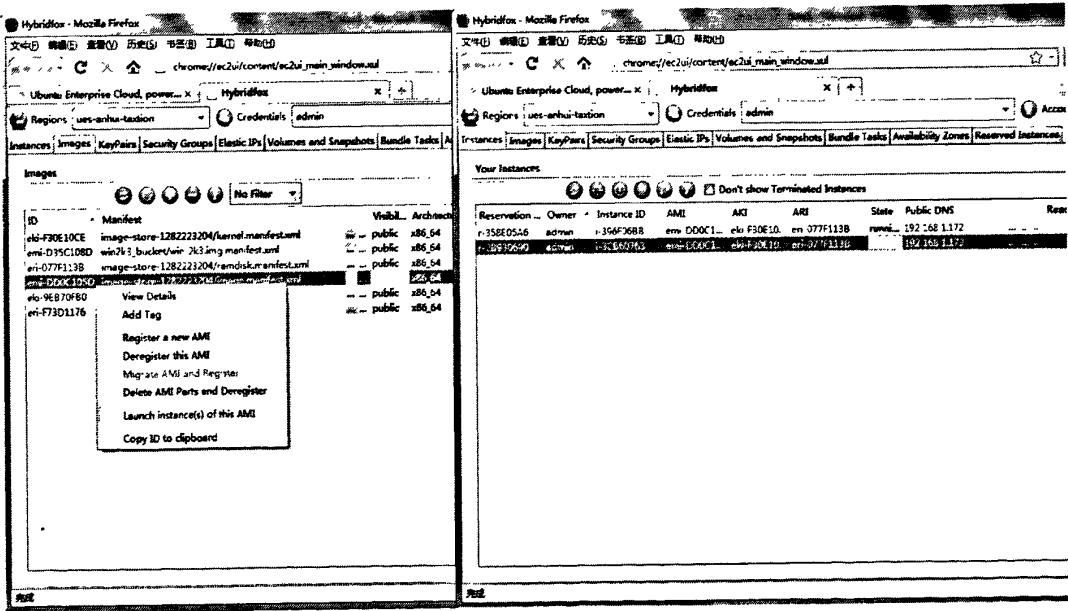


图 5-6 运行虚拟机实例及已运行的实例列表

Figure 5-6 Running virtual machine instances and a list of instances have been running

5. 从 Windows 操作系统远程登录到 UNIX 或 Linux 系统的虚拟机实例。

(1) 下载 PuTTY

([http://puttycn.googlecode.com/files/putty\\_0.60cn.zip](http://puttycn.googlecode.com/files/putty_0.60cn.zip)), 包括 PuTTY SSH 密钥生成工具 (puttygen.exe) 和客户端登录工具 (putty.exe), 解压缩到 “C:\Program Files\putty” (Hybridfox 的默认目录) 或其他目录。

(2)但是, PuTTY 不能识别生成的 PEM 类型的私钥, 需要使用 puttygen.exe 进行格式转换。

运行 puttygen.exe --> 在上方菜单选择 “转换 (Conversions)” --> “导入密钥 (Import Key)” ; 选择生成的 PEM 类型的私钥 (\*.PEM); 择 “保存私钥 (Save private key)” , 输入文件名即可生成 PuTTY 私钥文件 (\*.ppk)。

(3) 运行 putty.exe。选择 “会话 (Session)” , 输入 “主机名称或 IP (Host Name or IP)” ; 选择 “连接 (Connection) --> SSH --> 认证 (Auth)” , “浏览 (Browse)” 生成的 PuTTY 私钥文件 (\*.ppk)。

(4) 打开 (open), 使用用户名 “ubuntu” 登录, 成功登录后的窗口如图 5-7 所示。

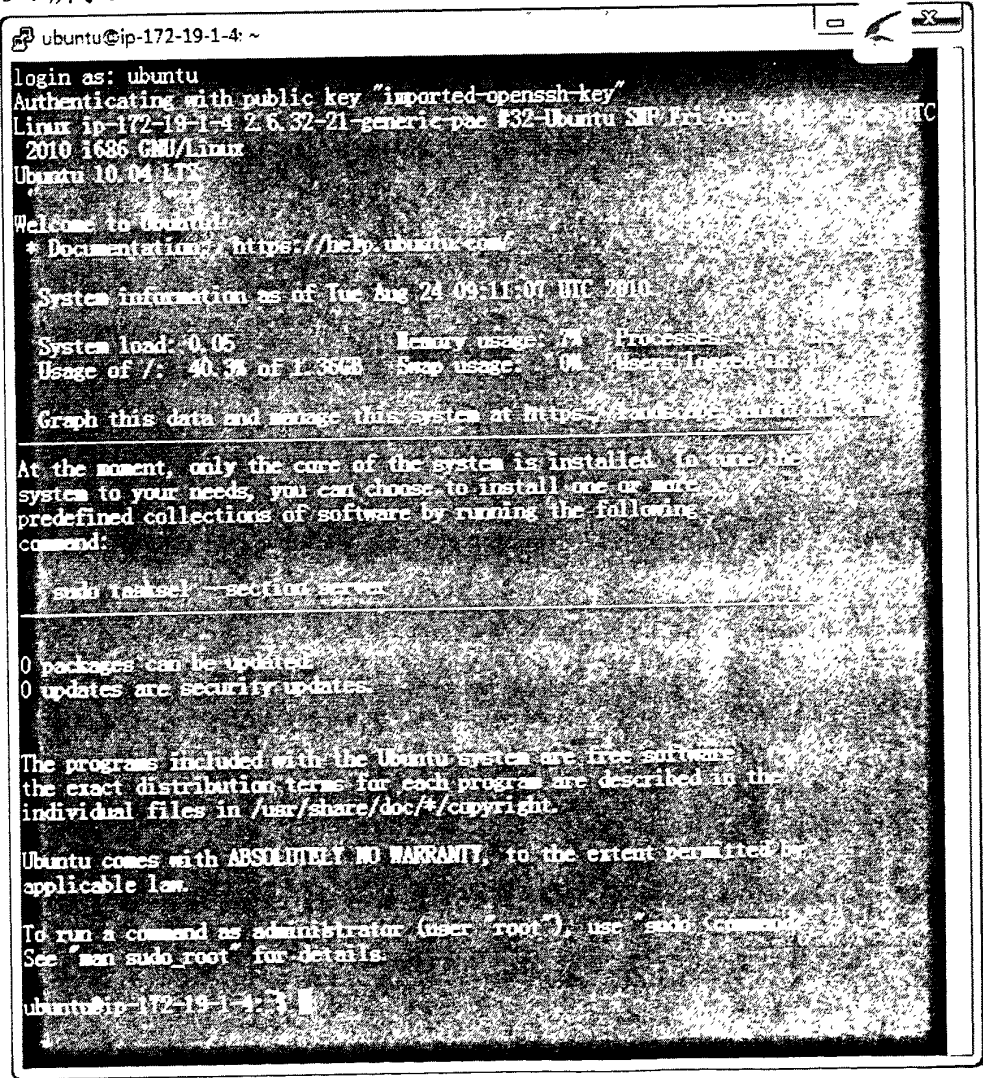


图 5-7 登录虚拟机实例  
Figure 5-7 Sign virtual machine instances



如果需要使用 HybridFox, 请选择右上角的“Tools”按钮, 在如图 5-8 所示界面中配置“SSH Command”(例如“c:\Putty\putty.exe”)、“SSH User”(例如“ubuntu”)等参数。否则在如图 5-6 所示界面中选择“Connect To Instance”, 会提示使用用户名“ubuntu”登录比用“root”登录更好一些, 然后自动关闭窗口。

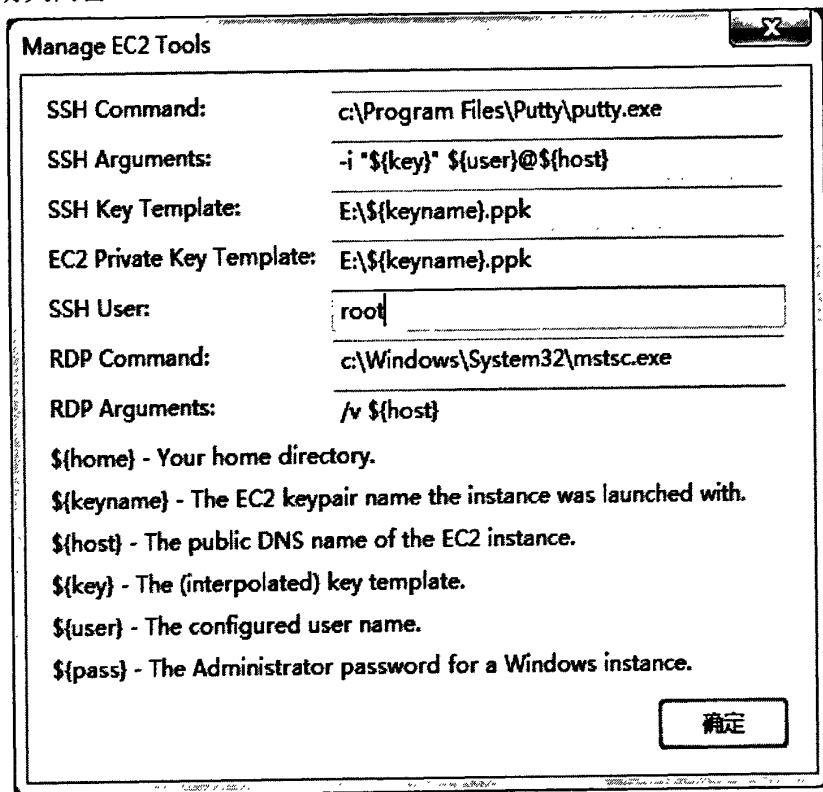


图 5-8 配置 EC2 工具

Figure 5-8 manage EC2 tools

## 5.4 部署应用

在本地完成系统开发、测试之后即可部署到私有云或公共云中, 整个部署过程相当于把应用系统迁移到云环境中。

部署 Web 应用系统到云中的一般过程是:

1. 从一个预配置虚拟机镜像库中分别选择作为负载均衡服务器、应用服务器和数据库服务器设备的镜像。
2. 在选择镜像的虚拟机实例中安装配置必要的组件, 并打包为自定义镜像。
3. 将数据库迁移到云中。
4. 更新本地应用系统配置, 使用云端数据库, 验证应用系统是否在本机可以正常运行。

5. 选择呈现并部署各层镜像的配置模式，将应用系统打包，并将它部署到云端应用服务器（虚拟机实例）。

6. 验证应用程序是否可以在云中运行。

本研究中云计算环境由一台IBM刀片服务器和两台PC机构成。具体部署过程是在启动的虚拟机实例中（如图5-7所示）下载并安装、配置JAVA、Weblogic Server for Linux, 通过Weblogic控制台将已生成的应用系统的ear包上传并部署，测试能够正常运行后使用euca命令行工具构建新的镜像（AMI）上传并注册。

## 5.5 性能测试

### 5.5.1 响应速度测试

Apache JMeter 是Apache组织的开放源代码项目，是100%纯java桌面应用程序，基于它的图形界面可以方便的进行压力测试和性能测量。它可以测试包括基于静态和动态资源程序的性能，例如Java Servlets、Jsp、Java 对象、静态文件、数据库、FTP 服务器等。JMeter可以用于在服务器，网络或对象模拟大负载来测试它们提供服务的承受能力或者分析不同压力条件下的总体性能。

本部署中系统响应速度测试的方法是利用JMeter工具分别在局域网环境和云环境中进行两次测试，局域网环境配置为一台IBM刀片服务器（Intel Xeon 8核2.5GHz，4GB内存，Windows2003 Server），云环境为局域网内的一台IBM刀片服务器（Intel Xeon 8核2.5GHz，4GB内存，Ubuntu10.04 UES）和一台PC（AMD Athlon II X4 640 3.0GHz，4GB内存，Ubuntu10.04 UES Node），虚拟机类型（VM Types）为ml.small（1CPUs，1G内存，30GB硬盘），2个虚拟机实例。JMeter参数设置包括线程数为500，Ramp-Up Period为0，循环次数为2，模拟出500个并发用户发出1000个取样请求。两次测试所得数据对比如表5-2和图5-9、图5-10所示。

指标	#Samples	Average	Median	90%Line	Min	Max	Error%	Throughput	KB/sec
第一次	1000	137	90	339	2	509	2.90%	1245.3/sec	2621.7
第二次	1000	29	11	88	2	212	0.00%	2288.3/sec	4815.8

表 5-2 第一、二次测试聚合报告  
Table 5-2 The First & Second Test Aggregate Report

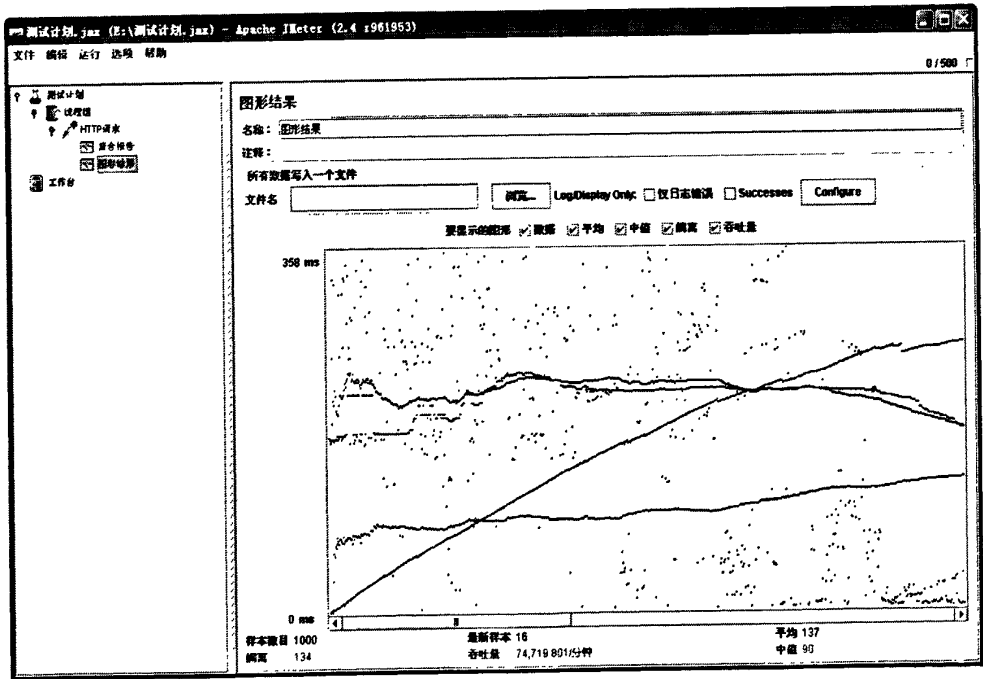


图5-9 第一次测试图形结果  
Figure 5-9 The First Test Graph Results

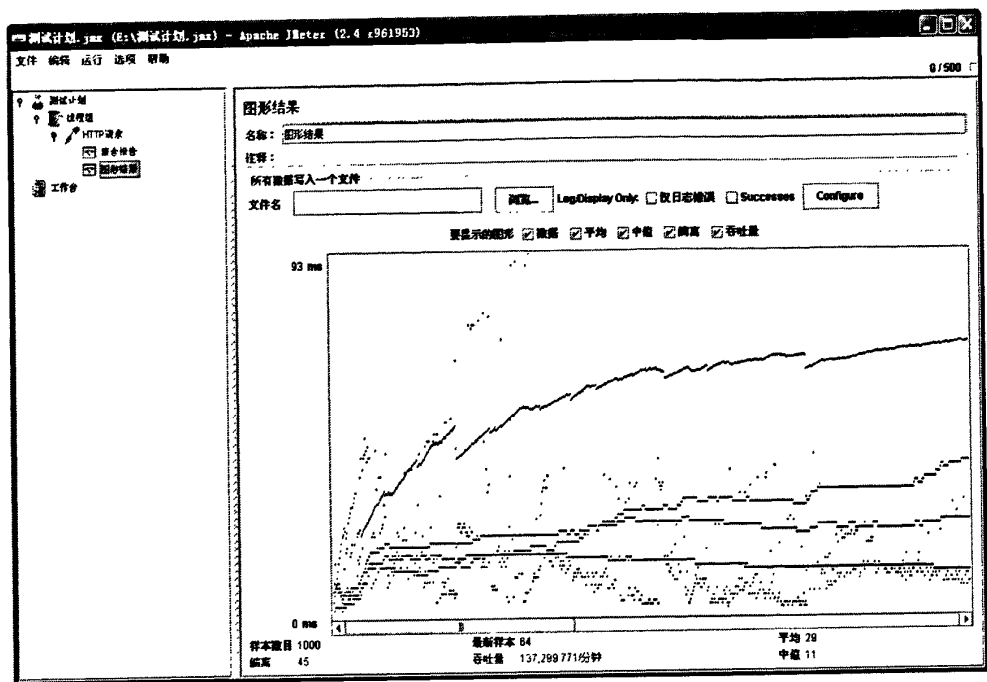


图5-10 第二次测试图形结果

Figure 5-10 The Second Test Graph Results

聚合报告显示，在第一次测试发出的1000个请求中，平均响应时间（Average）为137毫秒，50% 请求的响应时间（Median）不高于90毫秒，90% 请求的响应时间（90% Line）不高于339毫秒，最小响应时间（Min）为2毫秒，最大响应时间（Max）为509毫秒，发生错误比例（Error%）为2.9%，吞吐量（Throughput）为1245.3/秒，；数据量（KB/Sec）为2621.7KB/秒。在第二次测试发出的1000个请求中，不仅未出现错误请求，而且各项指标都明显优于第一次测试。

图形结果显示，在第一次测试发送到服务器的1000个请求中，服务器响应最后一个请求的时间（最新样本）为16毫秒，服务器每分钟处理的请求数（吞吐量）为74719，每个请求的平均响应时间（平均值）为137，有一半的请求响应时间低于而另一半高于90毫秒，响应时间变化、离散程度（偏离）为134；在第二次测试发送到服务器的1000个请求中，服务器响应最后一个请求的时间（最新样本）为64毫秒，服务器每分钟处理的请求数（吞吐量）为137299，每个请求的平均响应时间（平均值）为29，有一半的请求响应时间低于而另一半高于11毫秒，响应时间变化、离散程度（偏离）为45。

两次测试结果对比表明，基于云计算平台的应用系统具有良好的弹性扩展

能力，只要启动足够的虚拟机实例，可以完全满足访问量急剧增加的情形。

5.5.2 数据读写测试

在税收征管档案电子化过程中，上传、存储、查询资源是日常管理中的主要功能，由于部分资源所占空间较大，如何大文件读写效率也是难点。本研究在部署过程中，以运行在云环境中的MongoDB和局域网环境中的Oracle作为数据服务器分别进行图片或视频文件数据上传和读取测试。测试方法是选择3个大小不同的文件，分别上传、读取5次，测试结果(如表5-2所示)表明存储到基于MongoDB GridFS(map/reduce操作会并行运行在所有的shards上)建立的文件存储系统比存储到Oracle数据库的效率要高，文件数据量越大，这种优势越明显。

文件数据量	本地平均执行时间（毫秒）		云环境平均执行时间（毫秒）	
	写入	读取	写入	读取
314KB	47	31	0	0
5MB	1875	563	63	12
48MB	13117	5703	610	109
300MB	-	-	3391	522

表5-3 数据读写测试结果

Table 5-3 Test reading and writing data results

## 第六章 总结与展望

### 6.1 总结

主要的研究成果和创新可以概括为以下三点：

1. 通过对云计算理论的学习和私有云部署应用的实践，对当前主流的云计算的平台有了一定的研究。高效、经济的云计算平台将为构建电子政务起到助推作用，可以有效降低运营成本、提高设备利用率。同时，作为掌握着国民经济运行数据的政府部门，私有云将是首要选择。

2. 进一步研究了开源的Eucalyptus系统和分布式非关系数据库MongoDB，并将二者结合起来在高并发、大数据量交互的应用系统开发方面做了尝试，初步掌握了在云计算环境下的软件开发技术，进一步理解了软件工程理论。

3. 针对当前纳税服务体系存在的问题进行研究，通过云计算平台下的系统开发论证了纳税服务云平台建设的可行性和必要性，为将来纳税服务、税收核心业务的云平台建设总结和积累了经验。

由于时间和软、硬件设施等条件所限，本次研究还存在一些不足，主要表现在：系统还需要进一步完善；系统的数据存储采用开源的数据库并且数据量测试仅为GB级，无法和商业数据库在TB或PB级数据量上比较。但是，随着国税部门应用系统整合需求和程度的不断加大以及本人的不懈努力，一定能够在云计算理论研究和应用系统开发方面取得更多成果。

### 6.2 展望

就目前而言，云计算分布式软件开发缺乏开发方法学理论和易用、高效的开发框架支持，云存储服务也缺乏理论体系和开发框架支持。云计算时代的软件系统开发技术和理论体系还需要创新、发展和积累<sup>[16]</sup>（如图6-1）。然而，在Google、亚马逊、微软、IBM、SUN、Oracle、VMware等诸多大型企业的支持和努力下，云计算技术正在不断飞速发展，已经进入很多不同的领域，为科学计

算、电子商务等诸多领域带来革命性的变化，并且有着十分广阔和美好的应用前景。

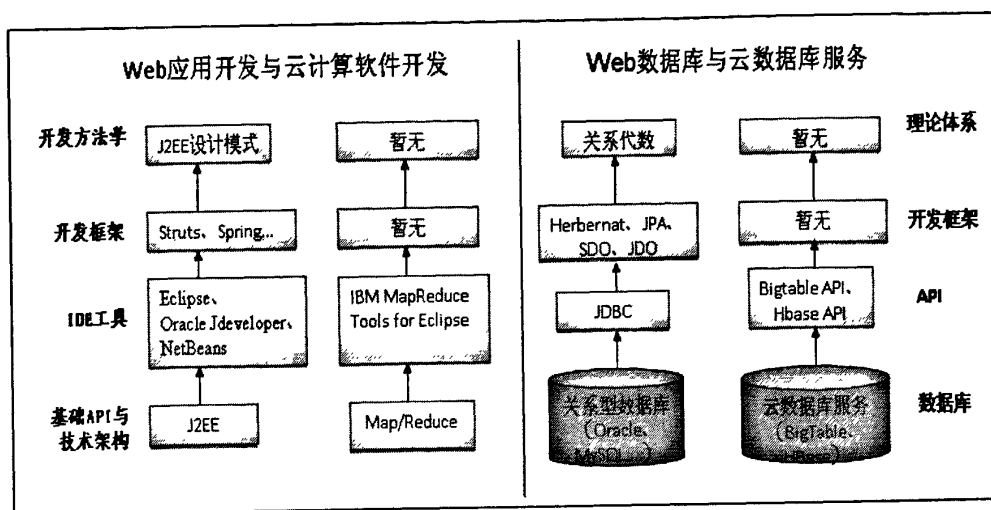


图 6-1 现有软件开发、数据库与云计算相关技术比较

Figure 6-1 existing software development, database, compared with the cloud computing related technologies

目前的纳税服务平台可以实现基础设施即服务的目标，而在此基础上的分布式应用开发、部署（软件即服务）还有待进一步研究。随着云计算软件开发、存储技术的不断发展，丰富的应用功能能够快捷地开发、部署到纳税服务云平台之中，满足纳税人多元化服务要求，税务机关的服务水平将实现质的飞跃。基于云计算的全国纳税服务平台的构建必将推动税收信息化水平的跨越式发展，成为我国税收信息化发展史上的里程碑。

## 参考文献

- [1] 马奎升. 纳税服务的思考与实践 [DB/OL]. (2008-06-08).  
[http://www.wfds.gov.cn/zjwfds/hxfz/11yt/200810/t20081008\\_2324.html](http://www.wfds.gov.cn/zjwfds/hxfz/11yt/200810/t20081008_2324.html).
- [2] 国家税务总局教材编写组. 税收基础知识 [M]. 北京: 中国财政经济出版社, 2009.
- [3] 陈涛. 云计算理论及技术研究 [J]. 重庆交通大学学报 (社科版), 2009, 9(4): 104-106.
- [4] 刘鹏程, 陈榕. 面向云计算的虚拟机动态迁移框架 [J]. 计算机工程, 2010, 36(5): 37-39.
- [5] 英特尔开源软件技术中心, 复旦大学并行处理研究所. 系统虚拟化: 原理与实现 [M]. 北京: 清华大学出版社, 2009.
- [6] 胡炜, 钟卫连. 浅谈云计算的网络安全问题 [J]. 中国科技信息, 2008, (23): 109.
- [7] 邓倩妮, 陈全. 云计算及其关键技术 [J]. 高性能计算发展与应用, 2009, 26(1): 2-6.
- [8] 李永先, 栾旭伦, 李森森. 云计算技术在图书馆中的应用探讨 [J]. 江西图书馆学刊, 2009, 39(1): 105-106.
- [9] 史田伟, 魏志强, 贾东宁, 杨永全. 基于云计算的智能家庭研究 [J]. 中国电子商情: 通信市场, 2009, (6): 232-236.
- [10] 百度百科. 云计算 [DB/OL]. <http://baike.baidu.com/view/1316082.htm>.
- [11] 陈德录, 吴泓. 云计算在税收服务平台中的应用研究. 电脑知识与技术, 2010, 6(21): 5815-5817.
- [12] CSDN. 张亚勤: 云计算的第一步是虚拟化 [DB/OL]. (2010-5-21).  
<http://news.csdn.net/a/20100521/218491.html>.
- [13] Ubuntu. Ubuntu Documentation [DB/OL]. <https://help.ubuntu.com/10.04/>.



- [14] Ubuntu. ElasticFox[DB/OL]. <https://help.ubuntu.com/community/UEC/ElasticFox>.
- [15] Eucalyptus. Eucalyptus Community[DB/OL]. <http://open.eucalyptus.com/wiki>.
- [16] 应时. 云计算时代的软件开发技术及其应用[DB/OL]. <http://www.chinacloud.cn/show.aspx?id=3086&cid=58>.
- [17] 史佩昌, 王怀民, 蒋杰, 卢凯. 面向云计算的网络化平台研究与实现[J]. 计算机工程与科学, 2009, (S1): 253-256.
- [18] 郭本俊, 王鹏, 陈高云, 黄健. 基于 MPI 的云计算模型[J]. 计算机工程, 2009, 35(24): 84-86.
- [19] 杨刚, 随玉磊. 面向云计算平台自适应资源监测方法[J]. 计算机工程与应用, 2009, 45(29): 14-17.
- [20] 陈康, 郑纬民. 云计算: 系统实例与研究现状[J]. 软件学报, 2009, 20(5): 1337-1348.
- [21] 王鹏, 董静宜. 一种云计算架构的实现方法研[J]. 计算机工程与科学, 2009, 31(A1): 11-13.
- [22] 曾诚, 李兵, 何克清. 云计算的栈模型研究[J]. 微电子学与计算机, 2009, 26(8): 22-24, 27.
- [23] 孙瑞锋, 赵政文. 基于云计算的资源调度策略[J]. 航空计算技术, 2010, 40(3): 103-105.
- [24] 左利云, 吴良海. 基于云计算的数据库查询调度算法[J]. 郑州大学学报(工学版) 2010, 31(4): 65-68.
- [25] 李伯虎等. 一种基于云计算理念的网络化建模与仿真平台——“云仿真平台”[J]. 系统仿真学报, 2009, 21(17): 5292-5299.
- [26] 刘鹏. 云计算——将计算变成水和电[J]. 中国计算机学会通讯, 2009, 5(10): 49-54.
- [27] 王龙, 万振凯. 基于服务架构的云计算研究及其实现[J]. 计算机与数字工程, 2009, 37(7): 88-91.
- [28] 袁国骏. 浅谈云计算及其发展应用[J]. 实验室科学, 2009 (2): 103-105.

- [29]赵玉国. 基于云计算的 WebGIS 系统构架研究[J]. 西昌学院学报·自然科学版, 2010, 24(2):45-47, 59
- [30]张敏, 陈云海. 虚拟化技术在新一代云计算数据中心的应用研究[J]. 广东通信技术, 2009, (5):35-39
- [31]黎春兰, 邓仲华. 论云计算的价值[J]. 图书与情报, 2009, 4:42-46.
- [32] 吕元智. 基于云计算的电子政务信息资源共享系统建设研究[J]. 信息系统, 2009, (9): 106-109.
- [33] 杨剑, 胡新平, 董建成. 基于云的区域电子健康档案存储和交换架构[J]. 医学信息学杂志, 2010, 31(3): 10-13.
- [34]陈谷. IBM 云计算和政府云平台[J]. 程序员, 2008, (11):56-57
- [35]高宏卿, 汪浩. 基于云存储的教学资源整合研究与实现[J]. 现代教育技术, 2010, 20(3):97-101.
- [36]Daniel J A. Data management in the cloud: limitations and opportunities. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering[J], 2009, 32(1):3-12.
- [37] Hayes B. Cloud computing[J]. Commun ACM, 2008, 51(7): 9-11.
- [38]James Murty. Programming Amazon Web Services[M]. US: O'Reilly Media, 2008.
- [39]Roger Jennings. Cloud Computing with the Windows Azure Platform[M]. US: Wiley Publishing, 2009.
- [40]吴吉义, 平玲娣, 潘雪增, 李卓. 云计算: 从概念到平台[J]. 电信科学, 2009, (12):23-30.
- [41]钱文静, 邓仲华. 云计算与信息资源共享管理[J]. 图书与情报, 2009, (4):47-60.
- [42]赵俊华, 文福拴, 薛禹胜, 林振智. 云计算:构建未来电力系统的核心计算平台[J]. 电力系统自动化, 2010, 34(15):1-8.
- [43]汪建, 方洪鹰. 云计算与无线局域网安全研究[J]. 重庆师范大学学报(自然科学版), 2010, 27(3):64-68.
- [44]陈康. 云计算后台大规模数据处理技术探讨[J]. 电信工程技术与标准

化, 2009, (11):12-16.

[45]赵华茗, 李春旺, 李宇, 周强. 云计算及其应用的开源实现研究[J]. 现代图书情报技术, 2009, (9):1-6.

[46]石磊, 邹德清, 金海. Xen 虚拟化技术[M]. 武汉: 华中科技大学出版社, 2009.

[47]王飞, 蔡建勋. 纳税服务理论研究[M]. 上海: 上海人民出版社, 2007.

[48]刘鹏. 云计算[M]. 北京: 电子工业出版社, 2010.

[49]王鹏. 走近云计算[M]. 北京: 人民邮电出版社, 2009.

[50]王鹏. 云计算的关键技术与应用实例[M]. 北京: 人民邮电出版社, 2010.

[51]CSDN. 中移动发布基于开源技术的大云平台系统[DB/OL]. (2010-5-21).

<http://news.csdn.net/a/20100521/218500.html>.

[52]赵永新, 我国首个自主开发云计算操作系统年内面世[N]. 人民日报, 2010-8-20(1).

[53]忆菲. 亚马逊正式推云计算服务数据将永久存储[N]. 电子资讯时报, 2008—8—28(A07).

[54]CNET 科技资讯网. 美奥巴马政府公布云计算计划 拟降低基础建设成本[DB/OL]. (2009-9-17). <http://www.cnetnews.com.cn/2009/0917/1469571.shtml>.

[55]左永刚, 云计算从概念神坛走向行业应用[N]. 中国高新技术产业导报, 2010-5-31(C1).

[56]陈骞. 国内云计算发展现状简述[DB/OL]. (2009-12-18). <http://www.istis.sh.cn/list/list.aspx?id=6373>.

[57]将您的 Linux 应用程序迁移到 Amazon 云[DB/OL]. (2010-8-2). <http://www.ibm.com/developerworks/cn/linux/1-migrate2cloud-1/>.

## 致 谢

在本人的学位论文完成之际，我首先要感谢安徽大学对我的教育和培养。在“至诚至坚，博学笃行”的校训精神的指引下，潜心耕耘、励志践行的师生们形成的浓厚的科研、学习氛围给我留下深刻印象，并将激励我在今后的工作、学习中取得更加优异的成绩。

感谢我的导师刘锋教授。刘教授对我的课题从选题、研究到完成给予了大量的建议、指导和帮助，付出了大量劳动。导师渊博的知识、严谨的治学态度、高度的责任心、和蔼可亲的学者风度将对我今后的工作和学习产生深远而积极的影响。

感谢软件学院的胡学联副教授。他在学术上一丝不苟，生活中平易近人。我感谢他把我领入到云计算这个研究领域之中，并给了我无私的帮助和教导。

感谢软件学院的郑蕊老师，有了她的严格的纪律要求和日常管理中的辛勤工作，我们才能取得更好的学习成果。

感谢软件学院2007届工程硕士班的同学胡贤德、张蕊、肖汉周、盛魁、吴元君、周翔宇、唐立，你们在学习、生活上给我很大的帮助，你们勤奋、主动的学习态度深深地影响了我。虽然在一起共同学习、生活的日子不多，但是我收获了一段人生最美好的回忆。非常感谢你们给予我的帮助，祝福你们在人生道路上不断取得成功。

感谢我所在工作单位的领导对我参加学位教育的关心和支持！

最后我要感谢我的妻子和女儿，没有她们的默默的付出，没有她们的理解和支持，我就不会在工作中取得优异成绩，更不可能顺利地完成学业。

陈德录

二〇一〇年九月二十日

## 攻读学位期间发表的学术论文目录

已发表的论文:

[1]陈德录,刘锋.基于B/S的网络考试系统高并发性能优化研究与实现.电脑知识与技术,2010,6(14):3658-3659.

[2]陈德录,吴泓.云计算在税收服务平台中的应用研究.电脑知识与技术,2010,6(21):5815-5817.

# 基于云计算的纳税服务平台的研究与实现

作者：[陈德录](#)  
学位授予单位：[安徽大学](#)

本文链接：[http://d.wanfangdata.com.cn/Thesis\\_Y1975524.aspx](http://d.wanfangdata.com.cn/Thesis_Y1975524.aspx)