



本科毕业设计（论文）

数据订阅系统的设计与实现

Design and Implement of Data Subscription System

学 院：____软件学院____

专 业：____软件工程____

学生姓名：____XXX____

学 号：____XXX____

指导教师：____XXX____

北京交通大学

2016 年 5 月

学士论文版权使用授权书

本学士论文作者完全了解北京交通大学有关保留、使用学士论文的规定。特授权北京交通大学可以将学士论文的全部或部分内容编入有关数据库进行检索，提供阅览服务，并采用影印、缩印或扫描等复制手段保存、汇编以供查阅和借阅。

（保密的学位论文在解密后适用本授权说明）

学位论文作者签名：

指导教师签名：

签字日期： 年 月 日

签字日期： 年 月 日

中文摘要

摘要：近年来，O2O 互联网经济模式发展迅速，尤其是国内各大互联网公司之间的 O2O 大战使得该经济模式尤为引入注目。这种快速增长促使公司其他业务部门对于 POI 数据的需求越来越频繁。公司其他业务部门在申请 POI 数据时存在的主要问题是：现有的申请业务系统管理松散，工作效率低。该问题的主要原因是：由于业务沟通几乎全是通过电子邮件方式传达信息，灵活性低，沟通效率低、二是限于业务本身的要求，需要更新数据，重新走一遍流程过于繁琐，浪费时间和精力，三是生产数据环节中 RD 手工处理脚本，易出错，业务执行效率低。为解决以上问题，需要设计一套业务流程优化和自动化的优选方案，在保证服务提供的数据准确无误的前提下，减少整个业务流程执行的平均耗时，最大实现业务自动化运行。

本文的目标是基于对现有业务执行效率影响要素与技术途径的分析，提出业务流程优化和自动化的方案，并给出该方案的主要模块的设计与实现。本人主要工作是通过调研分析并完成 workflow 管理系统实现、数据安全传输、消息通讯以及业务自动化实现的技术途径优选以及总体方案的设计与实现。

通过量化手段对优选方案进行验证评估，通过对业务流程的优化以及自动实现，预估计缩短 60% 的业务平均运行耗时。提升员工工作效率，对于数据信息化管理中数据流向的管理更加有力。

本文的主要结构为：首先介绍研究背景、研究目标与意义等，分析对于业务执行效率影响要素与技术途径优选，之后对业务流程本身进行建模分析，最终给出方案中主要模块的具体设计与实现。

关键词：数据订阅自动化；工作流；数据安全；即时通信；

ABSTRACT

In recent years, the Internet O2O economy has developed rapidly, especially O2O war between major domestic Internet companies so that the economic model is particularly compelling. This rapid growth prompted the company to other business sectors POI data demand for more and more frequent. The main problem the company's other business units in the application POI data exists is: the existing business application systems management loose, low efficiency. The main reason for this problem is: Since almost all business communication is to communicate by e-mail information, flexibility, low efficiency of communication, the second is limited to the requirements of the business itself, the data needs to be updated, re-take over process is too cumbersome, and a waste of time energy, the three links in the production data RD manual processing script, error-prone, low operational efficiency. To solve the above problem, we need to design a business process optimization and automation of the preferred embodiment, to ensure the accuracy of the data service provider premise of reducing the average time to perform the entire business process, to achieve maximum automation of business operation.

Objective of this paper is the efficiency of the existing business impact analysis features and technical approach proposed business process optimization and automation solutions, and design and implementation of main modules of the program are given based. It mainly works through research and analysis and complete workflow management system implementation, design and implementation of technical means secure transmission of data, messages and communications to achieve business automation and preferred overall program.

Verified by quantifying the assessment of the preferred embodiments of means, through the optimization of business processes and automatically pre-estimated reductions of 60 percent of the average running time-consuming business. Enhance employee productivity, data for information management in data flow management more powerful.

The main structure of this paper is: introduce research background, objective and significance, operational efficiency analysis for influencing factors and technical approach is preferable, after the business process modeling and analysis itself, and ultimately give a specific design of the main modules and achieve.

KEYWORDS: data subscription automation; workflow; data security; instant messaging;

目 录

中文摘要.....	I
ABSTRACT.....	II
目 录.....	III
1 引言.....	1
1.1 课题研究背景.....	1
1.2 研究目标与意义.....	1
1.3 研究实现难点及关键点.....	2
1.4 交付物.....	2
1.5 专有名词介绍.....	2
1.6 论文组织结构.....	3
2 理论基础与技术途径的优选.....	4
2.1 业务流程分析方法概述.....	4
2.1.1 范围分析法.....	4
2.1.4 观点分析法.....	4
2.2 数据传输安全的技术途径优选.....	6
2.3.1 HTTP.....	6
2.3.2 HTTPS.....	7
2.4 消息通讯的技术途径优选.....	9
2.4.1 电子邮件.....	9
2.4.2 WebSocket.....	10
2.5 业务自动化实现的技术途径优选.....	10
2.6 系统总体方案设计优选.....	11
2.7 本章小结.....	12
3 业务流程分析优化与需求定义.....	13
3.1 数据订阅系统介绍.....	13
3.2 As-Is 模型建立.....	14
3.2.1 业务流程的利益相关者分析.....	14
3.2.2 基于服务提升和办公自动化的业务边界范围分析.....	15
3.2.3 当前业务 As-Is 模型建立.....	16
3.2.4 业务流程分析.....	18
3.3 业务流程优化.....	22
3.3.1 业务基础环境假设.....	22
3.3.2 单个环节改进方案.....	22
3.4 未来业务 To-Be 模型的建立.....	27
3.5 系统需求分析.....	27
3.5.1 系统主要功能模块划分.....	28
3.5.2 系统用例图.....	28

3.6 本章小结.....	29
4 数据订阅系统整体设计与实现.....	30
4.1 WFMC workflow 管理系统体系架构	30
4.2 workflow 管理系统参考模型.....	31
4.3 数据订阅系统体系结构及系统组件组成.....	32
4.4 workflow 管理系统各组件设计.....	32
4.4.1 workflow 过程定义.....	33
4.4.2 workflow 引擎.....	33
4.4.3 workflow 监控.....	35
4.5 数据订阅系统角色图.....	35
4.6 数据订阅系统主要模块功能划分.....	36
4.6.1 系统整体功能框图介绍.....	37
4.7 系统表单模块实现.....	37
4.7.1 角色权限管理模块.....	37
4.7.2 数据申请表单模块.....	38
4.7.3 定时器模块.....	39
4.8 WEBSOCKET 即时通信设计与实现.....	41
4.8.1 数据库 Message 表设计.....	42
4.8.2 WebSocket 即时通信业务类图设计	44
4.8.3 在线通信设计方案.....	44
4.8.4 离线通信设计方案.....	45
4.8.5 WebSocket 即时通信业务时序图	46
4.9 本章小结.....	46
5 系统分析及优越性验证.....	48
5.1 系统业务流程优化 KPI 分析.....	48
5.2 系统业务升级的管理提升.....	48
5.3 业务执行效率的提升.....	49
5.4 系统优化实现的风险评估.....	50
5.4 本章小结.....	50
6 结论.....	51
参考文献.....	52
致 谢.....	53
附 录.....	54

1 引言

本章主要分别介绍课题的来源与背景，课题研究目标及投入研发的意义，研究的难点，论文的内容以及组织结构。

1.1 课题研究背景

本课题源于本人在百度时代网络技术（北京）有限公司实习的研发项目，即数据订阅系统的设计与实现。

本人所在的实习团队为 POI 数据团队，主要负责管理和维护地图所有的 POI 数据。现阶段，由于 O2O 互联网经济模式的崛起，公司其他业务部门对 POI 数据的需求也越来越频繁，POI 数据信息化管理中的数据订阅业务自然成为团队中业务发展的核心关键^[1]。

现阶段团队内的 POI 数据订阅业务流程存在诸多发展问题：简单的电子邮件消息通讯模式越来越不能满足频繁的业务需求；依赖电子邮件进行申请需求提交以及审批批复，导致申请内容及条件等不清晰而无形中增大沟通成本，影响业务整体执行效率；没有有效的角色和权限管理机制，安全性低；业务流程执行状态不可见，影响申请人员的工作安排，无法了解历史操作管理，对于数据流向管理造成管理上的漏洞；生产数据环节需要 RD 手动执行脚本传参执行，容易出现错误，并且重复的操作也浪费和分散 RD 大量的时间以及精力。

通过以上分析挖掘的问题，我们不难看出，现有的数据订阅业务流程松散，人为因素多，且效率低。而一个合理的数据订阅系统首先应当有一个强大的表单系统，能够满足各种需求及条件的自定义配置；强大的定时器方案推动高度实现自动化的可能，高度实现自动化运行是结构化工作流系统的唯一优化目标方向；有效的角色及权限机制，传输数据安全的机制等是系统正常运行的基础保证。完善以上几点才能更好地保证数据订阅业务流程高效运行，大幅提高服务质量与能力。

由于数据申请的业务量不断增大。为了能够更好地满足大量业务对部门高效数据订阅系统的需求，需要对现有业务流程进行优化和最大自动化实现。

1.2 研究目标与意义

本论文研究的数据订阅系统是以高度自动化且安全的方式提供服务。接入公司内部的用户认证平台，方便用户统一登录管理，良好的角色和权限的配置管理，构建表单系

系统将申请元数据结构化，提供自定义频率（定时器）运行及推送服务，数据传输过程安全加密，自动调度脚本生产数据，流程状态及监控管理。本平台的研究与实现，将有利于其他业务部门更方便快速地申请所需 POI 数据，第一时间获取业务流程状态，大幅减少流程执行过程中人为因素的影响，导致业务执行失败或者系统出现不可恢复的灾难情况。也有利于团队更好把握数据流向的管理。

1.3 研究实现难点及关键点

本文的难点在于如何有效完成数据订阅系统的优化以及自动化实现。主要体现在以下几点：

- (1) 对于申请表单系统的结构化设计，能够满足所有的数据申请需求（即元数据条件），实现自定义频率（定时器）运行并完成推送。
- (2) As-Is 模型及未来业务的 To-Be 模型建立
- (3) 统一管理该业务的通讯消息，提高沟通效率和灵活性。
- (4) 角色与权限管理，流程的状态及数据流向管理。
- (5) 保证数据传输安全。

关键点在于两个方面，第一个是表单的结构化设计，该表单系统需要满足其他部门的员工各样的申请需求。使得使用表单系统提交的申请都能符合 POI 团队的要求，并且也方便了其他部门员工。表单的结构化设计当中还包括最重要的定时器设计。另一个方面在于统一该业务的通讯消息，提高沟通效率和灵活性。

1.4 交付物

- (1) 数据订阅系统的架构设计
- (2) 数据订阅系统平台框架及代码

1.5 专有名词介绍

(1) Point Of Interest，又称作兴趣点，表示电子地图上的某个地标，用以标示出该地所处的政府部门、商业机构、旅游景点、名胜古迹、交通设施等处所。

1.6 论文组织结构

论文第一章是引言部分，分别对课题研究背景、研究目标及意义、论文研究难点及关键点的介绍、交付物进行详细阐述。第二章是理论基础及技术背景的介绍，主要对论文中涉及的理论知识及技术的说明，并介绍技术途径优选的过程。第三章是对业务流程进行分析，通过应用二章中列举的各种分析法进行认真的分析，深入挖掘根本问题的源头，为下一步系统设计与实现作好准备对象工作。第四章是业务流程优化及自动化的实现过程，即为系统整体设计与实现。分别介绍了系统主要角色，系统的角色及权限模块设计，系统的表单结构化模块设计以及定时器模块设计，最后着重介绍了 Socket. IO 库的在本项目的使用，给出在线及离线两种通信设计方案。第五章是项目分析和优越性验证，利用量化手段验证新模型的优越性。第六章是全文最后的总结部分，笔者对整篇论文进行小结，提出自己的见解以及对未来系统发展的展望。

2 理论基础与技术途径的优选

本章主要是基于对现有业务流程执行效率影响要素与技术途径的分析，通过对各项技术与方法的优劣分析，介绍了本数据订阅系统的理论基础和技术背景，为本平台的设计与实现提供强有力的理论及技术支持。确定问题的解决思路与途径，并体现技术途径的优选的过程。

2.1 业务流程分析方法概述

本节主要通过对业务流程分析中，常用的方法进行必要的概述，例如范围分析法、观点分析法。下面将详细介绍各分析方法的使用过程。

2.1.1 范围分析法

针对某个具体的项目，范围分析指的是为了完成某些特性及功能，给出该项目应该完成的工作内容。通常情况下，对于项目的实施之前，都需要对项目所涉及到的范围进行明确，这样可以更好地保证项目进度及质量，而且可以提高项目的开展的可靠性等。

范围分析法通过明确项目的范围，为之后的项目进展提供必要的保证，主要可分为以下几步：

- (1) 对利益相关者进行适当调研，帮助确定项目的制约条件，限制，不同利益相关者关注的地方。
- (2) 描绘出清晰的业务流程图，并且对项目边界加以明确

2.1.4 观点分析法

观点分析法是一种有效的分析工具，可以帮助开发团队更好确认系统需求，明确各项需求的优先级，观点分析的产出物一般是观点机构图或者是观点分析表。在图表当中对不同系统利益相关者的观点进行归纳及分析，进一步对系统的需求有了更加明确的认识^[5]。观点分析的好处主要分为下面几点：

- (1) 图表的形式能够更加层此化的对需求进行展示及分析

（2） 图表形式可以帮助更好地挖掘客户需求的根源

（3） 有利于项目进展，允许整个团队共享信息的方法，能够达到团队的一种共同认知。

2.2 workflow 管理系统实现途径的优选

workflow 技术源于办公自动化领域，是以业务流程集成为方向，基于模型的业务过程自动化实现的技术。workflow 管理系统是 workflow 技术为主导的软件管理系统。workflow 管理系统存在的目的在于为实现某个业务执行目标，需要使用计算机在多个用户角色中按照某种既定的规则进行文档，信息或者任务自动传递。workflow 技术为自动化管理软件的创新提供了新思路，成功把 workflow 应用到自动化管理系统可以帮助提高办事效率，降低生产成本，增强对用户的服务能力，有效管理业务流程等目的。

通过对 workflow 技术的大量文献检索发现不同 workflow 系统所采用的任务传递途径的不同，主要可以分为以下四种：

结构化 workflow 管理系统：结构化 workflow 管理系统指的是在业务执行过程当中该项业务需要经常重复操作，并且是依照某一个既定的规则进行的业务活动。制定该 workflow 的过程中需要对业务流程本身作大量的分析。许多办公事务入审批、文档处理等都属于这类管理系统范畴。

基于文件分发的 workflow 管理系统：该系统的显著特点在于通过以电子文档形式对文件进行分发处理操作，目的是方便该业务系统当中的各个角色能够很好地完成处理与审阅的工作。这样的 workflow 系统常常被描述为一个序列执行操作过程，各个环节当中都有相应的待处理的数据对象。

基于消息通讯的 workflow 管理系统：该类系统主要是通过以电子邮件的通讯方式来完成过程当中业务执行结果消息的传递，数据分发以及事件的通知。缺点在于这样会使得整个业务系统处理一种相对松散的管理模式下。优点在乎充分利用电子邮件系统在广域环境条件下的数据转发能力。

基于群件的工作流管理系统：该类管理系统主要依赖于系统本身自有的应用基础结构，包括有消息传递，安全保护，目录功能以及文档服务等。其本身已经是一个完成的应用开发环境。

笔者通过对上面四大类型工作流的分析，发现数据订阅系统本身属于一个审批类管理系统，业务操作重复性高，有一个完好的既定规则分发处理机制，流程处理步骤明晰，我们可以清楚看到基于结构化的工作流系统是符合我们系统本身的要求。数据申请属于批量重复业务行为，可以通过结构化对业务流程进行完整的定义。

2.2 数据传输安全的技术途径优选

保护 POI 数据安全对于 POI 团队来说至关重要，对于数据订阅业务流程当中数据传输安全是一个非常关键的问题。本节主要对 Web 管理系统的传输协议进行研究与分析，主要分析 HTTP 协议与 HTTPS 协议之间的区别和优缺点。最后给出一个对于保证数据订阅服务中数据安全的最优技术途径。

2.3.1 HTTP

HTTP 是用于 Web 平台进行传输 HTML 文件以及其他多用途互联网邮件扩展资源的协议约定。HTTP 在客户端以及服务端之间遵循请求-应答模式。默认的 HTTP 请求目的端口为 80 端口^[11]。常用客户端工具为浏览器，又常被称作用户代理程序。应答服务上存储有各种各样的互联网资源，例如文件以及图像，音视频资源等。使用 HTTP 协议进行规范双方传输数据包时存在一个明显的缺陷是明文传输。明文传输的结果就是使得双方通讯信息暴露在 Internet 上，下图 2-1 给出了 HTTP 请求与应答的简单关系图。



图 2-1 HTTP 请求与应答

报文是 HTTP 通信当中的基本传输单位。报文是由八位字节流组成，并且通过 HTTP 通信传输。且 HTTP 报文本身是由多行（CR+LF 作换行符）数据所构成的字符串文本块。客户端叫作请求报文，服务端叫作响应报文。下图 2-2 描述了 HTTP 报文结构，图 2-3 描述了请求报文与响应报文的详细结构图。



图 2-2 HTTP 报文结构

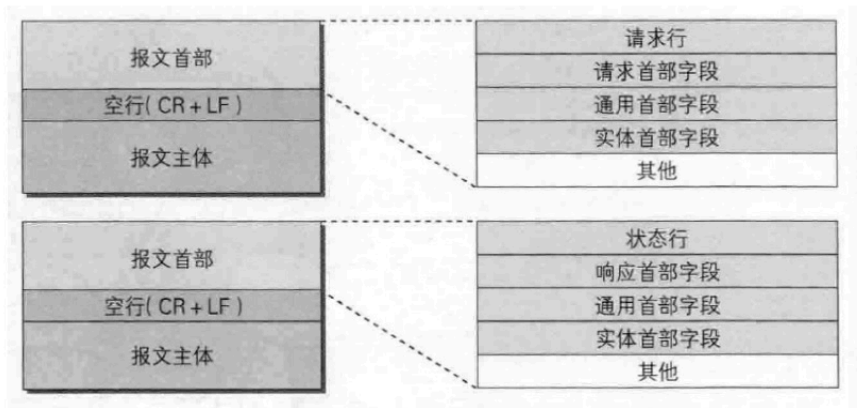


图 2-3 请求报文与响应报文详细结构图

2. 3. 2 HTTPS

超文本传输安全协议 **Hypertext Transfer Protocol Secure**，缩写为 **HTTPS**，也被称为 **HTTP over TLS**，**HTTP over SSL** 或 **HTTP Secure**，是一种网络安全传输协议^[12]。使用 HTTPS 协议可以保护传输内容得到加密，具有身份认证机制保证，并且确保数据传输过程保持完整性。HTTPS 通常被认为是 HTTP 协议与 SSL/TLS 协议的组合。SSL（Secure Sockets Layer）中文叫作“安全套接字层”，为传输层加密协议。下面给出 HTTP 和 TLS 在 TCP/IP 协议族中的位置以及 TLS 协议的组成结构，如下图 2-4 TLS 协议格式：

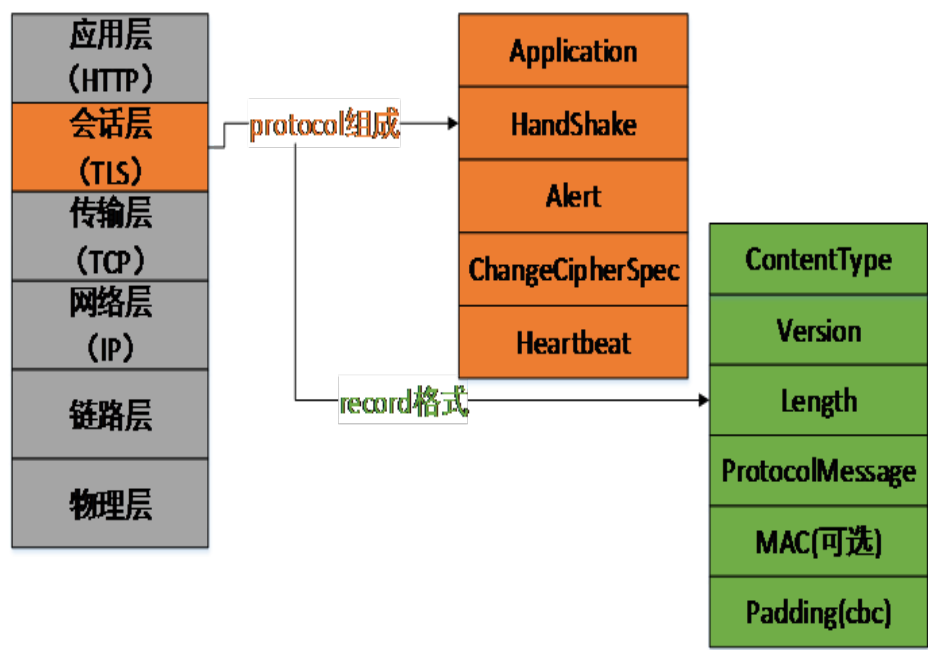


图 2-4 TLS 协议格式

下面将给出 HTTPS 中关于协议握手部分的介绍，见下图 2-5 HTTPS 运作图：

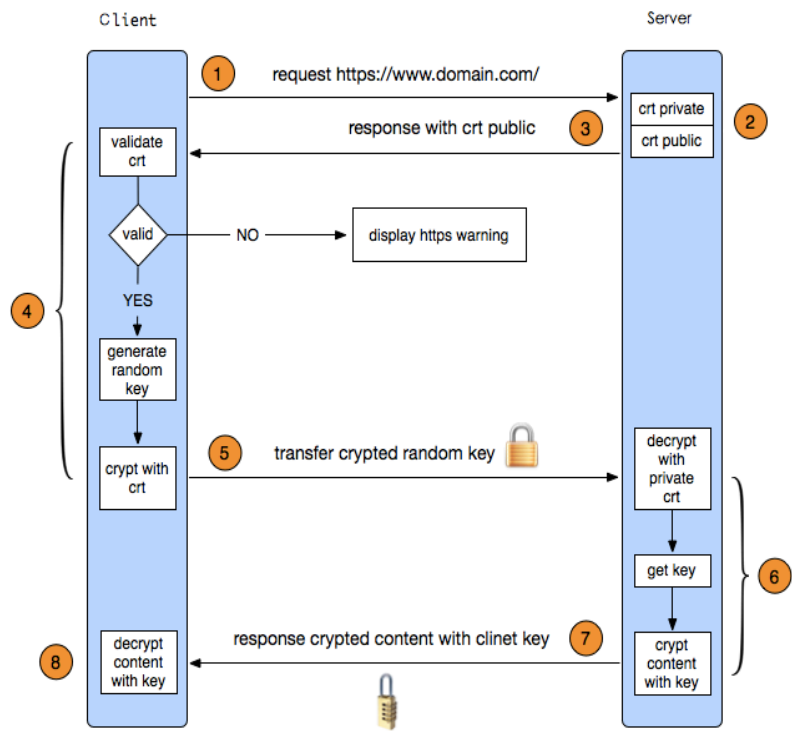


图 2-5 HTTPS 运作图

上图 2-5 主要描述了 HTTPS 握手中密钥交换的过程,是为后续内容加密作准备工作,通常这回引入对称加密与非对称加密的比较,对称内容加密的强度虽然非常大,但是存在一个严重问题是无法安全生成及管理密钥。非对称加密主要用于密钥交换,浏览器和服务器协商过程中都使用非对称加密算法进行协商对称密钥,用于后续的传输内容加密。

由于 POI 数据为机密数据,为了确保数据传输安全,通过笔者对 HTTP 及 HTTPS 协议的研究过程得出 HTTPS 加密传输是本系统安全传输的关键,因此确立数据安全传输的方案是基于 HTTPS 来实现。

2.4 消息通讯的技术途径优选

本系统现有的消息通讯是依赖于电子邮件通信方式进行一切业务沟通与数据来往媒介,下面将对电子邮件通讯技术以及 WebSocket 全双工通信技术的分析比对,优选适合于本系统的消息通讯技术途径。

2.4.1 电子邮件

电子邮件,英文简称 email,中文简称电邮,是指一种由寄件人将文本信息发送给一个人或者抄送给多个人的信息交换方式。早期的电子邮件是需要发件人和收件人同时在线,类似现在的即时通信工作模式。目前的电子邮件系统主要是以存储和转发为运行模型的基础,电子邮件起初是用 ASCII 编码邮件内容,随后技术不断更新,越来越多的扩展开始加入,最后产生了多用途互联网邮件扩展。下图 2-6 给出了一般寄送电邮的流程:

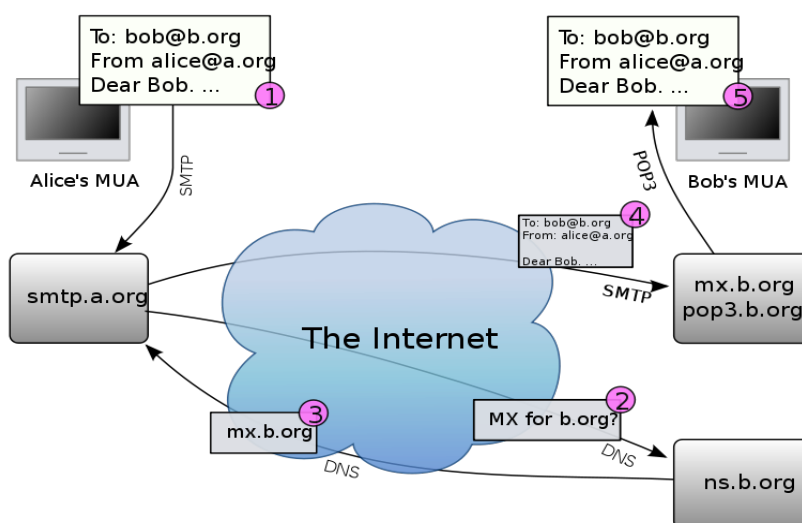


图 2-6 寄送电邮流程图

2.4.2 WebSocket

WebSocket 是自 HTML5 开始提供的一种在单个 TCP 连接上进行全双工通讯的协议。通常在 WebSocket API 中，浏览器和服务器需要首选在 HTTP 协议层面上完成一个握手的动作，然后浏览器与服务器之间才能形成一条 WebSocket 基于 TCP 传输层的通信通道。两者可以开始进行数据的相互传送。

在 WebSocket 未出现以前，许多网站为实现伪全双工通信几乎都是通过使用 ajax 轮询来实现的，通常是在特定的时间间隔内，由客户端向服务器不断发出特定的 HTTP 请求，而后由服务器将新数据返回给对应的客户端。然而此种传统的模式带来很严重的问题是，客户端需要不停地向服务器发出 HTTP 请求，又由于 HTTP 请求报文里头会携带大量的头部信息，所以当请求里面包含的数据仅仅只是几 kb 的传输量的话，服务器端的带宽以及资源就被大量浪费了。虽然后来的 Comet 技术，在长连接技术应用使用上改善了很多，但根本仍然会大量消耗服务器带宽和资源。

WebSocket 实现原理主要如下：请求头部中 Sec-WebSocket-Key 是随机产生的，服务端使用该数据构造一个 SHA-1 的信息摘要。之后将 Sec-WebSocket-Key 加上字符串 258EAF5-E914-47DA-95CA-C5AB0DC85B11。使用 SHA-1 加密，再进行 Base-64 编码，将结果作为 Sec-WebSocket-Accept 头部字段的值返回给客户端。

通过以上对电子邮件和 WebSocket 两种通讯技术的分析对比可知，当前业务系统当中应用的电子邮件通讯机制使得系统的管理松散，电子邮件系统中通常混有各样的不同的信息内容，对于邮件用户来说通常处理邮件要事不是特别即时，因此对于需要高效率的业务来说电子邮件通讯方式无疑是一个很大的阻碍。而 WebSocket 是基于 Web API 提供服务的，因此对于整个数据订阅系统来说，全双工通信的 WebSocket 技术的引入必定大幅提升沟通效率，提升业务流程整体的工作效率。

2.5 业务自动化实现的技术途径优选

现有业务流程因为存在人为因素过多的问题，导致数据订阅业务整体执行效率过低，对于业务自动化实现的技术途径优选显得尤为重要。基于系统业务本身的特点，现提出基于 Crontab 技术可以实现定时器管理业务运行^[9]。

使用 crontab 命令用于设置定时器周期性执行的指令工具。该命令是从标准输入设备读取命令，并将输入指令存储到文件 crontab 中，供系统读取以及读取。一般来说，Crontab 程序储存的命令是由系统守护进程发现并执行的，通常是在系统后台执行，系统会每隔一分钟检查是否有预定的作业需要完成，通常其被称为 crontab jobs^[10]。

Crontab 操作符号介绍如下：

- (1) 逗号（,）隔开，例如：“2, 4, 6, 9”
- (2) 连词符（-）指定值的范围，例如：‘2-7’，其等同于‘2, 3, 4, 5, 6, 7’
- (3) 星号（*）表示任何值，例如：在分钟域中，星号表示每一分钟。

时间格式说明如下图 2-12：

```
# 文件格式说明
# —分钟 (0 - 59)
# | —小时 (0 - 23)
# | | —日 (1 - 31)
# | | | —月 (1 - 12)
# | | | | —星期 (0 - 7) (星期日=0或7)
# | | | | |
# * * * * * 被执行的命令
```

图 2-12 crontab 指令时间格式

2.6 系统总体方案设计优选

通过研究与理论的分析，提出最终几种方案如下：

(1) 申请开设一个专用邮箱，用于公司其他业务部门发送提交申请数据元数据信息，使用时间及使用用途等。指定 RD 轮流排班处理，执行脚本，生产数据，打包作为附件并回复申请邮件对方。开发成本几乎为零，统一管理规范有效，但是人为因素影响依然存在，许多业务部门对申请邮件所需的信息，内容要求以及格式并不了解，因而导致在沟通交流当中浪费多余重复的精力。

(2) 开发一个普通的申请系统，由公司其他部门直接填写，提交，实现统一管理申请工单，再由部门项目经理确认审批，RD 着手执行脚本，生产数据，再交付对方业务部门。开发成本低，周期短，但依然不能有效解决现有问题。

(3) 借鉴 workflow 系统的主要思想，结构化的 workflow 系统作为模板参考，添加 HTTPS 化加密安全传输，有完善的角色及权限管理，表单结构化处理，数据定时执行并完成推送，使用 Socket.IO 库实现系统即时通信，流程状态及监控管理，优化前端设计及交互。开发成本较高，周期长，充分考虑长远利益，大幅提高服务质量，使得原有业务流程能够得到大幅度的优化及高度自动化。使得系统的服务质量能够得到更好地提高。

认真分析完这三种方案，按第一个方案来看，这个方案确实是零开发成本，只需要向上面申请一个专用邮箱，就可以统一处理及回复申请邮件，但是无法避免中间因为各种不必要的沟通成本导致精力的浪费。而且大量人为因素仍然存在。第二个方案与第一个方案类似，开发一个简单的申请系统，可以使得表单结构化申请元数据，可以适当地减少沟通交流的精力浪费。第三个方案主要借鉴优秀的结构化的工作流系统的思想，表单结构化申请数据的元数据信息，实现完善一体的角色及权限管理，HTTPS 保护数据传输安全，有效的流程状态及监控管理，Socket. IO 库实现即时通信，脚本传参自动执行机制，专注的前端交互及设计帮助用户更好地享受本系统提供的服务。开发成本相比前两个方案都高，但是大幅度减少人为因素的影响，实现系统自动化运行处理。因此本论文选出的最优方案为第三个方案。

2.7 本章小结

本章作者主要通过基于现有对业务流程执行效率的影响因素以及技术途径进行分析并优选最佳方案及时。关于提出对适用本论文研究的基础理论及技术背景，主要涉及到几种常用于业务流程分析的方法：范围分析法，情境分析法，瓶颈分析法，观点分析法等分析问题及定位目标的技术手段。通过介绍工作流技术，以及分析市场上流行的几种工作流管理系统，并优选实现工作流管理系统的途径。为实现数据安全传输，分析了 HTTP 及 HTTPS 协议的工作原理及区别，而后又分析两种消息通讯方式：电子邮件与 WebSocket 协议，分析两者之间应用的优缺点，评估带来的收益，优选 WebSocket 用于实现系统的即时通信。最后提出 Crontab 技术用于解决和实现当前业务流程人为因素多，导致数据订阅业务执行效率低的问题，目标是实现业务高度自动化。在最后一节总结概括基于对理论基础及技术途径的优选，为进一步的研究奠定有力的基础。

3 业务流程分析优化与需求定义

笔者在本章节中首先介绍下本系统的主要内容，然后通过分析建立 As-Is 模型，进而对其进一步的分析，找出需要优化的环节与范围，并给出 To-Be 模型。分析业务流程中的诸多科学的分析办法，范围分析主要指的是在已知需要完成某些特性及功能的前提下，明确项目应当完成的工作内容，帮助确认实现的目标范围。观点分析提供了该系统中不同参与者对当前业务流程的看法与建议，根源分析方法通过结构化的方式处理问题，瓶颈分析与情境分析分别通过找出业务流程的瓶颈所在之处，以及考虑各种可能发生的结果，分析未来的可能发生时间的过程，通过考虑分析各种结果及影响，帮助决策作出更加明智的选择。

3.1 数据订阅系统介绍

数据订阅系统，Data Subscription System, 以下简称为 DS，是百度时代网络技术（北京）有限公司下百度地图部门为了提升内部 POI 数据申请服务效率而开发的数据申请信息化管理系统的名称。根本目的是通过对原有业务流程加以优化，使用 workflow 技术构建结构化 workflow 系统管理和支持数据订阅的服务，高度自动化实现，提高工作效率，降低数据传输安全，以便更好地为公司其他业务部门提供 POI 数据获取服务。下图 3-1 是百度地图 POI 团队业务整体框架图：



图 3-1 POI 团队业务整体框架图

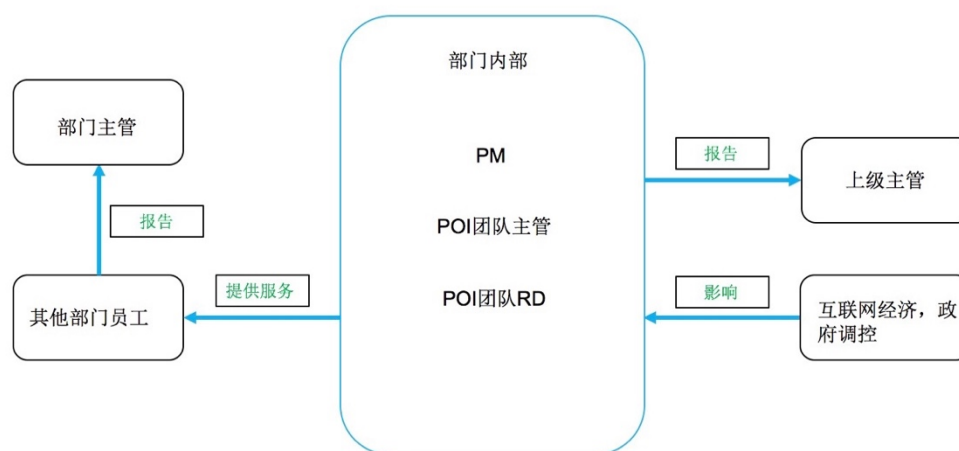
DS 系统主要涉及到的是 POI 数据管理方向，数据管理中包括权限管理环节，申请管理环节等。部门希望能够通过 DS 系统的上线，来解决目前 POI 数据申请业务当中存在的一些问题，为 POI 数据信息化管理奠定基础。DS 系统也将会帮助部门对 POI 数据申请业务持续以一个高效运行的服务提供，弃用基于电子邮件通讯方式以及手工处理工作的管理模式，最终达到真正实现系统全自动办公，并为部门内部对数据流向管理提供数据报告和分析。

3.2 As-Is 模型建立

本节笔者主要通过对部门数据订阅业务进行利益相关分析，找出优先级高的待解决的问题以及未来业务发展的目标。根据对本系统利益相关者进行调研和分析，提炼 KPI 需求点。接着根据上面分析结果，获得系统边界范围。并根据系统范围进行提炼现有订阅业务的 As-Is 模型。

3.2.1 业务流程的利益相关者分析

本节将通过对业务流程的利益相关者进行调研分析，帮助确定系统的关键点和重点亟需处理的问题。下图 3-2 描述了本系统的项目利益相关者之间的关系，其中主要的利益相关者分为部门内部和部门外部两个范围，其中部门内部包括：PM、POI 团队主管、POI 团队 RD。部门外部包括公司其他业务部门员工以及对应业务部门主管。互联网经济和政府调控会影响部门的决策。



笔者对以上的利益相关者进行了问答调研，并且对获得的结果进行研究分析，通过利益相关者目标矩阵分析方法，获得了利益相关者目标矩阵，给出每一种利益相关者提出的系统关键点。最后通过按优先级进行排序，确定本系统的实现目标和关键点。下表 3-3 为利益相关者目标矩阵表：

表 3-2 利益相关者目标矩阵表

角色/目标	降低成本	办公自动化	提升服务	降低失误
其他部门员工	*	****	****	****
其他部门主管	*	***	***	*
PM	**	**	***	***
POI 团队主管	**	****	***	***
POI 团队 RD	*	****	***	***
总计	7	17	16	14
优先级	4	1	2	3

上表 3-2 给出了四个系统目标是通过系统利益相关者调研筛选得到的重要性最高的，通过对表格的观察可以发现办公自动化以及提升服务能力的优先级是最高的，笔者也将这两点做点后续研究的重点。接下来对业务边界范围的界定也将根据上述得出的优先级结果进行研究。

3.2.2 基于服务提升和办公自动化的业务边界范围分析

基于上节对系统利益相关者分析获得的结果，本系统的主要关键点在于提升服务和实现办公自动化，笔者在下面给出了提升服务和办公自动化的业务边界范围表格，目的是为了体现出业务边界范围。如小表 3-3 所示：

表 3-3 业务边界范围表

提供者	输入	主要流程	结果输出
部门业务研发	员工填写申请数据	发放申请权限	数据结果

对 POI 数据的需求 的元数据信息及条件 部门主管级 POI 主管审批
数据生产及数据全部生成

为了能够更加清楚地看到系统边界与各个组成部分之间的紧密的关系和相互影响，如下图 3-3 所示：

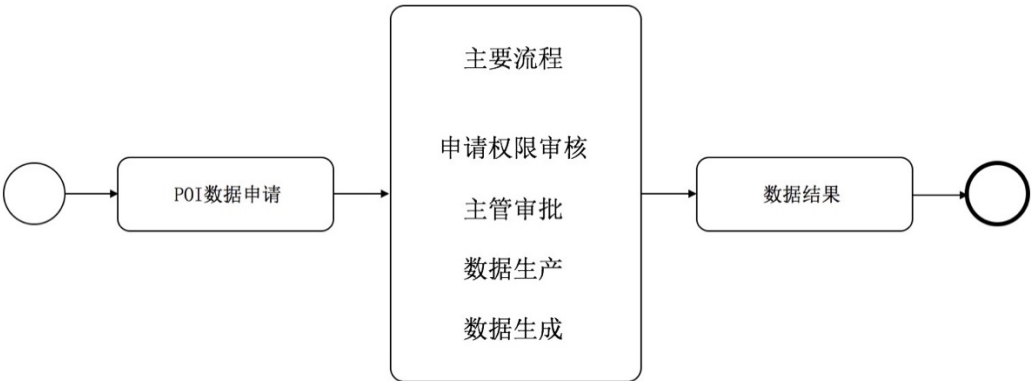


图 3-3 业务边界范围图

现有业务流程由于过分依赖电子邮件通讯方式和人工操作，其中数据生成过度依赖手工进行的，工作效率低。无法对数据进行评估报告上级主管。

3. 2. 3 当前业务 As-Is 模型建立

在上面的章节中通过对系统利益相关者的调研分析得到系统的边界范围，作者将在本节通过对系统目标、重点和关键点进行结合并生成业务申请流程图，最终确定本系统的范围。根据系统范围计算出现有业务的 As-Is 模型。下图 3-4 为现有业务的申请流程图：

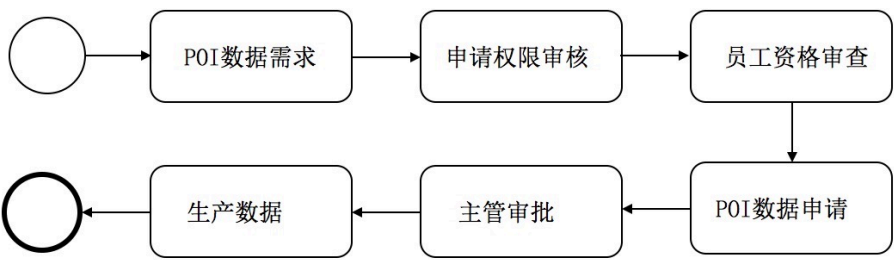


图 3-4 现有业务申请流程图

POI 团队为公司其他业务部门提供 POI 数据申请服务，当其他业务部门有 POI 数据的需求时，通过向 POI 团队的 PM 进行申请 POI 数据申请的权限，PM 对其资格进行审查，通过后再为其添加角色以及相应的权限，员工拿到权限后，进入数据申请系统填写 POI 数据申请表单并完成提交，申请完成后由该部门主管进行审批，该部门主管审批完成后，再由 POI 团队主管进行最后审批，审批通过后进入数据自动生产环节，最后完成数据结果交付。

根据以上多项的研究与分析，笔者通过制作得到部门现有申请业务的 As-Is 业务流程图，见下图 3-5：

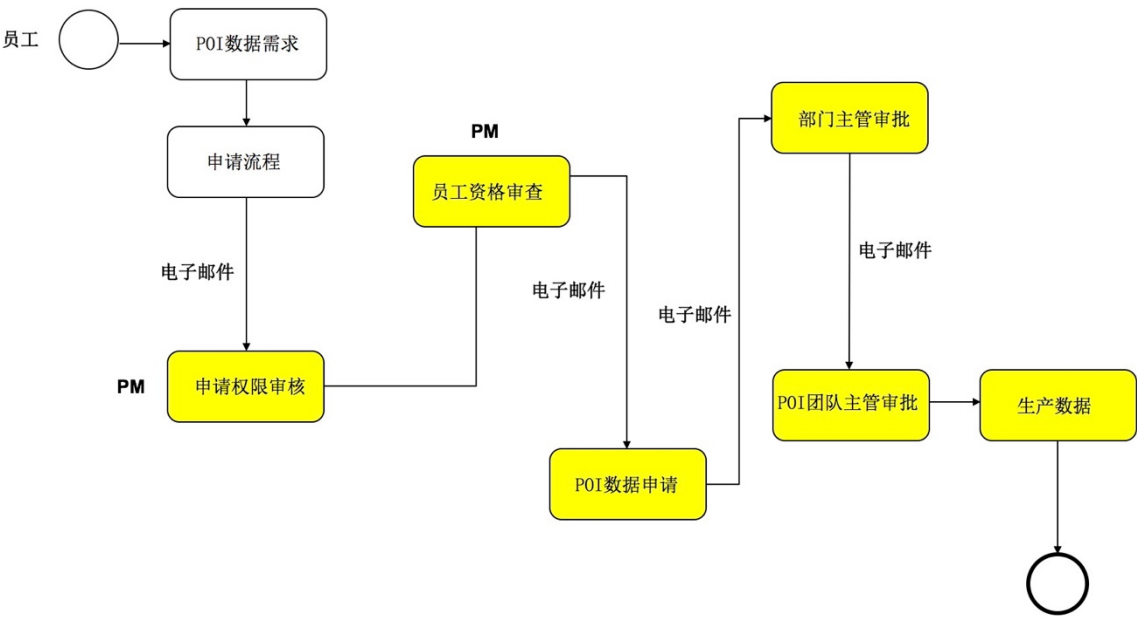


图 3-5 As-Is 整体业务流程图

3.2.4 业务流程分析

本小节主要使用本文中第二章介绍的用于业务流程的科学分析方法进行分析，范围分析主要对系统的范围进行分析并确定，观点分析提供了本系统中不同利益相关者的看法与建议，瓶颈分析指出了当前业务流程中存在的解决点及瓶颈处。最后综合以上的分析结果，生成数据订阅业务的鱼骨图，总结分析需要改善环节。

业务流程范围分析

本章节主要通过对现有的业务流程进行明确范围，通过对系统使用相关的用户进行适当调研，确定了本项目的一些制约条件。底下是笔者通过总结和收集调研意见，完成基本业务流程图的制作。

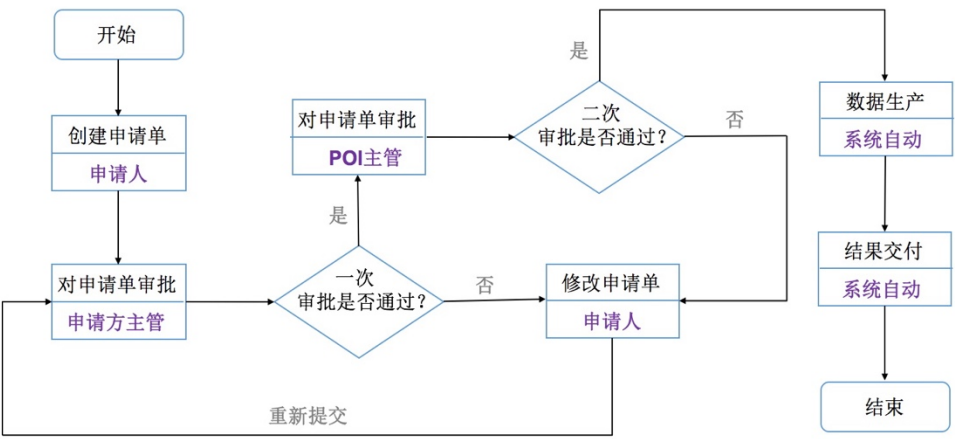


图 3-6 数据订阅系统的核心部分业务流程图

经过小组成员与公司其他业务部门的沟通，其对口人员期待有一个完整的系统，能够实现表单结构化数据元数据信息，包括复杂条件自由组合等形式。并且可以实时获知业务流程执行状态的情况。业务流程执行过程实现最大自动化优化。与 POI 项目经理沟通的结果是同样期待有一个完整的申请审批系统，审批流中需首先由对方部门所属上级主管审批通过后才流向本部门审批，同样需要严格把控数据传输安全，并且能够实现消息即时通知功能。与 RD 的沟通是希望能够实现脚本动态传参执行，避免 RD 因需要手动

调用脚本执行而浪费过多精力，使得工作效率降低。对三种利益相关者的调研沟通后，总结并详细规划处核心部分的流程图：如图 3-1 所示，由对方部门相关 RD 登录系统填写申请单并提交，后由该部门上级主管审批通过，之后再由 POI 项目主管审批，并进入自动生产环节，自动生成 map/reduce 脚本，从 Hadoop 集群上拉取数据。数据产出完毕，修改数据库表信息，并邮件及系统内部通知申请方数据准备就绪，可供拉取。

业务流程观点分析

上面通过对业务流程进行范围分析得出了整个业务系统的系统范围即从完成申请到最后数据产出并交付对方员工，下面将使用观点分析对整个业务流程进行认真调研分析。观点分析方法是一种有效的分析总结工具，主要目标是为了根据收集到的不同利益相关者提出的观点进行深入分析。目的是达到更好的收益效果。本项目的利益相关者分别为：公司其他业务部门员工，公司其他业务部门主管，POI 项目主管，POI 方 RD，接下来讲简单罗列各个利益相关方的意见与看法。最后以表格的形式列举出来。以便清晰地表达不同利益相关者的想法。

公司其他业务部门员工：期待有一个完整的系统，能够实现表单结构化数据元数据信息，包括复杂条件自由组合等形式。并且可以实时获知业务流程执行状态的情况。避免因不熟悉申请数据时所需要的必要内容的信息，导致在沟通上往复造成时间的浪费，业务流程状态的实时可见也能帮助我们预估数据的产出时间，有助于我们合理安排我们的工作。

公司其他业务部门主管：希望能够弃用电子邮件的通讯方式，使用即时通信来替代。避免过多邮件导致处理不及时，而且公司邮件系统也不是特别好用，希望能够弃用抄送审批回复的形式，有灵活的审批机制来取代更为合适。不但提高了效率，同样也是节省时间和金钱。

POI 团队主管：也是希望能够弃用电子邮件的方式进行沟通交流，因为我的邮箱里头每天有太多的邮件，如果没有及时处理的话，很容易漏掉这些需求邮件。因此隔离采用独立的系统较为合适。这样可以提高效率，当然也能够节省金钱和时间。

POI 团队 RD：希望能够自动配置脚本去 Hadoop 集群上拉数据，弃用手动调用脚本的

方式，避免因手动操作过程中因为拼写，漏写字段等常见错误，导致资源的浪费以及时间的浪费，这样也可以帮助我们能够将精力集中在重要的工作上。

综合上面的意见来看，作者经过适当提炼，将以上不同角色的观点进行分析和概括，如下表 3-2 所示：

表 3-2 观点分析概括表

根本目标：提高业务执行效率（自动化实现），降低执行成本	
公司其他业务部门员工	观点：减少以及避免不必要的沟通成本、提升工作效率 措施：表单结构化元数据信息、增加业务流程状态管理
公司其他业务部门主管	观点：提升业务流程执行效率 措施：弃用电子邮件通讯方式，建设统一通信系统。
POI 团队主管	观点：提升业务流程执行效率 措施：弃用电子邮件通讯方式，建设统一通信系统
POI 团队 RD	观点：提升业务效率，减少 RD 精力不必要的浪费 措施：动态配置脚本执行拉取数据

本章节主要是基于业务在按照最高效率运行的业务流程进行，各个业务流程均无失误出现，且各个业务单均能顺利完成。基于这种情况对其进行瓶颈分析，根据影响本业务流程效率响应的环节，作为本业务流程待改善的重点。

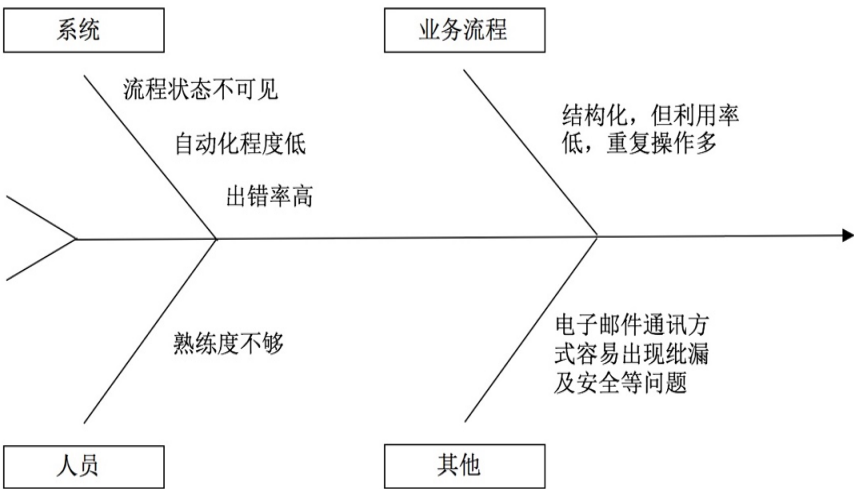


图 3-7 现有业务流程鱼骨图

通过下面表 3-3 的内容，描述在单位小时内现有业务流程内部各个不同的角色日常处理工作的耗时的情况，公司其他业务部门员工，公司其他业务部门主管，POI 团队主管，POI 团队 RD 的基本日常工作任务耗时，记为各个业务环节单位小时内业务处理的能力。通过对比，可以帮助我们清楚看见 POI 团队 RD 根据申请信息手动调用脚本跑数据环节是实现最高效率运行业务流程的瓶颈所在之处，也是解决优先级别最高的一项。业务流程原本遵循的是木桶效应（木桶效应还可称作短板理论，核心思想为：单位木桶的盛水量，实际上不取决于其桶的外围最高的那块木板，恰恰相反取决于其外围上最短的那块木板，又可称为短板效应）^[5]。因此抓住瓶颈的关键处，我们可以通过解决瓶颈出的处理能力来达到提高整个业务流程的执行效率的目的。

表 3-3 本系统各个角色每单业务流程处理用时表

角色	业务	业务能力
其他业务部门员工	邮件中填写申请信息	时间：15 分钟/单
其他业务部门主管	审批该部门申请邮件	时间：5 分钟/单
POI 团队主管	审批其他部门申请邮件	时间：20 分钟/单

POI 团队 RD

根据申请信息生产数据

时间：30 分钟/单

3.3 业务流程优化

数据订阅系统的业务操作属于高度流程化软件管理系统。在本文第三章部分，笔者通过运用科学的业务流程分析方法对数据订阅业务流程进行了深入细致地分析，发现了许多关键问题和优化的重点环节。因此在本章节中，笔者将尝试解决已经发现的问题并给出改进的建议。笔者在此通过模拟假设业务环境，假定一定的业务基础条件，应用之前分析得到的改进点对现有业务流程进行优化并生成未来业务的 To-Be 模型。

3.3.1 业务基础环境假设

笔者在这里给出数据订阅业务基础环境假设，下面章节的设计与实现将遵循以下假定基础环境：

- （1） 跨部门沟通关系融洽，遵循公司利益至上原则。
- （2） 在新系统未上线前，部门仍然按照现有业务流程完成每项 POI 数据申请业务，新系统的研发不影响现有业务的正常工作。
- （3） 社会经济模式以及政府调控没有发生特别大的变化。
- （4） POI 数据需求量保持平稳发展。

3.3.2 单个环节改进方案

在本节中，笔者将对前面分析得出的 As-Is 业务流程图进行分解并细化改进各个环节。最后将各个环节改进后的部分进行拼接，完成 To-Be 模型的构建工作。

- （1） 申请权限审核环节

申请权限审核环节的业务流程为：起始点通过电子邮件向 POI 团队 PM 说明关于申请 POI 数据的需求，POI 团队 PM 直接为之配置响应角色及权限。下图 3-8 为 As-Is 模型当中的身躯权限审核环节的流程圖：

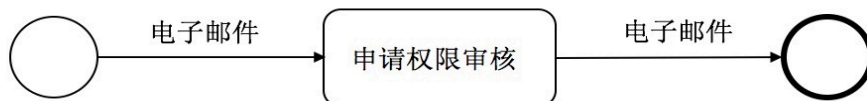


图 3-8 As-Is 模型申请权限审核环节的流程圖

依据前面的分析结果，下图 3-9 为笔者对申请权限审核环节进行流程优化，如图 3-10 所示：

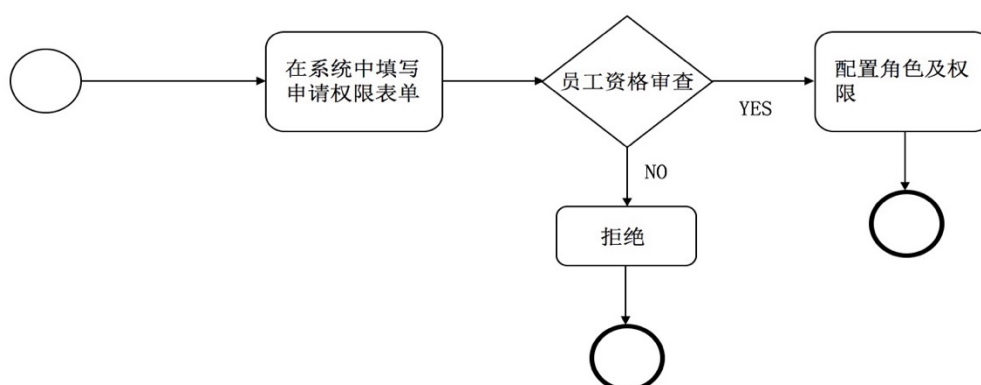


图 3-9 To-Be 模型申请权限审核环节流程图

POI 团队 PM 会手动对该员工的所在部门以及该员工进行资格审查,为保护 POI 数据的机密,PM 会向该部门主管进行询问是否为真正的需求,即为员工资格审查。如果是则为之配置角色及相应的权限,否则丢弃拒绝其申请权限请求,流程结束。

本环节的优化提升主要体现为下面两点:

(a) 弃用电子邮件通讯方式,通过表单结构化申请权限需求内容,实现流程自动化,提升业务管理能力。

(b) 增加员工资格审查判断功能,大大提升工作效率。

(2) POI 数据申请及审批环节

下图 3-10 为 As-Is 模型当中 POI 数据申请及审批环节的流程图,起始点通过使用电子邮件的通讯方式将 POI 数据需求信息,使用用途,数据字段信息,什么时间要,以及各种条件信息发送给 POI 主管和部门上级主管,主管通过电子邮件进行抄送回复审批。详情参见下图 3-10 所示:

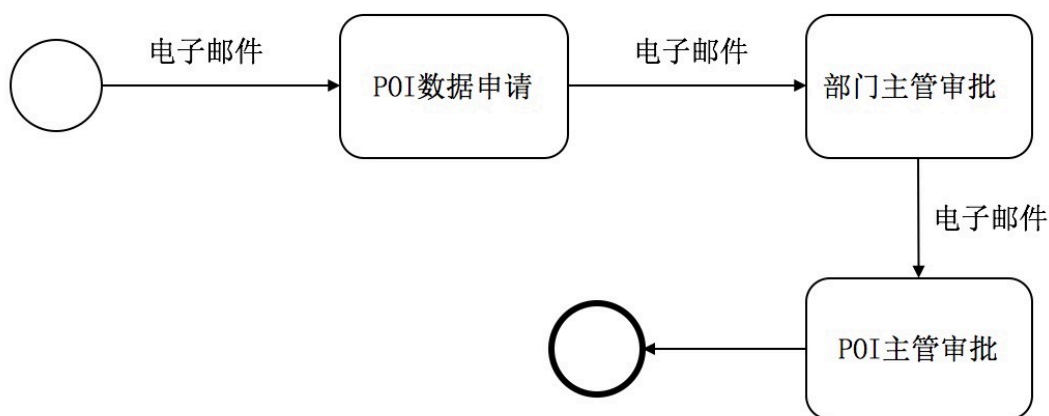


图 3-10 As-Is 模型 POI 数据申请及审批环节流程图

根据上面的分析结果，作者在此给出 POI 数据申请及审批环节的优化结果，如下图 3-11 所示：

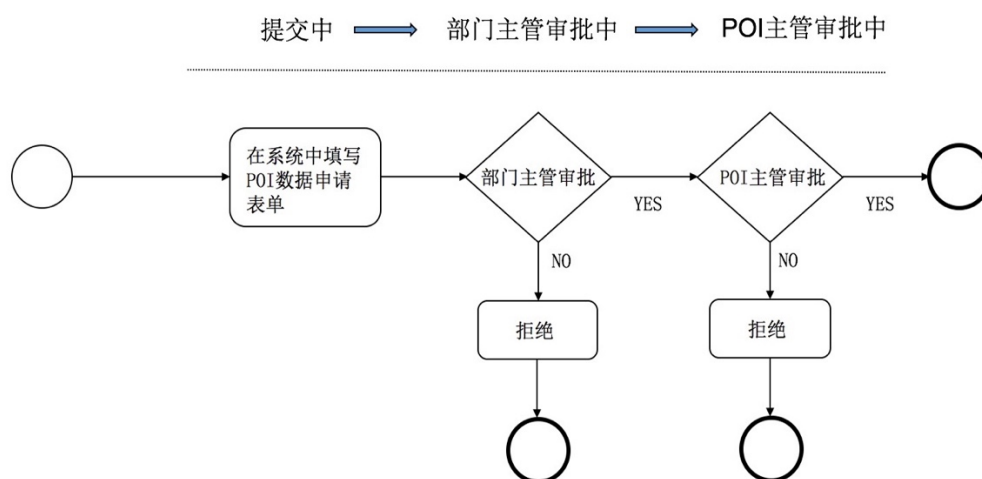


图 3-11 To-Be 模型 POI 数据申请及审批环节流程图

笔者对本环节的优化中，提升的方面主要体现在以下几点：

（a） 实现流程自动化，弃用电子邮件的通讯方式，使用表单系统进行申请及审批管理。

（b） 增加流程状态管理，帮助其他部门员工能够实时掌握业务运行的状态，帮助其做好工作安排，提高工作效率。

（3） 生产 POI 数据环节

下图 3-11 为 As-Is 模型当中生产 POI 数据环节的流程图，起始点通过使用电子邮件的通讯方式将 POI 数据需求信息，使用用途，数据字段信息，什么时间要，以及各种条件信息抄送给 POI 团队负责对外数据申请业务线的 RD，RD 通过分析邮件主体内容，提取申请数据的元数据信息及条件等信息，并调用编写 Map/Reduce 脚本去 Hadoop 集群拉取所需的数据，完成后通过电子邮件以附件形式将结果数据交付给对方员工。流程结束。详情参见下图 3-11 所示：

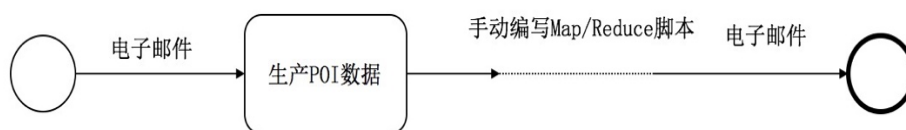


图 3-11 As-Is 模型生产 POI 数据环节流程图

结合本章中对流程自动化实现中 RD 提出的观点以及利益相关者分析的结果看，数据生产自动化是本系统的重中之重，避免手动编写脚本时输入数据元数据信息时出现拼写及漏写等人为因素非常明显的错误，进而导致系统工作不稳定，效率低。本小节中笔者对生产 POI 数据环节进行了较大的优化和改进，通过后端动态配置数据元数据信息并生成相应的 Map/Reduce 脚本，如果执行错误可由人工进行修改脚本并手动再执行。详情见下图 3-12 所示：

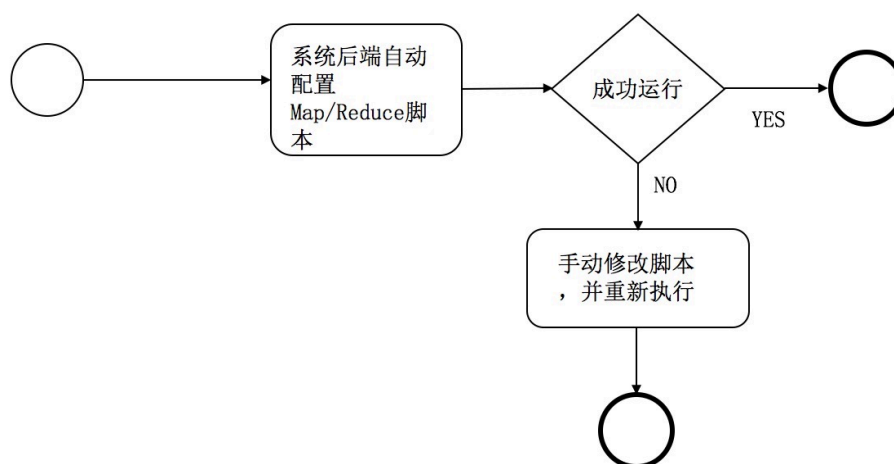


图 3-12 To-Be 模型生产 POI 数据环节流程图

本环节的优化当中，提升的主要方面有以下几点：

（a）实现数据生产自动化，省去 RD 手动编写 Map/Reduce 脚本，避免人为因素的影响导致系统服务不稳定。

(b) 弃用电子邮件的通讯方式，减少不必要沟通带来效率的降低。

3.4 未来业务 To-Be 模型的建立

笔者在对之前的分析总结出以下几个亟需解决的问题，包括有：电子邮件通讯方式沟通不灵活，效率低，管理混乱，邮件内容结构性差、主管审批耗时长、业务流程执行状态不可见，无法实时监控、出于业务需要不断更新数据，反复申请耗时耗力、数据流向管理能力弱，无法进行客观统计。通过分析和总结，得出亟需解决的两个问题：业务自动化及服务能力的提升。下面将给出基于对现有业务分析和改进得到数据订阅系统的 To-Be 模型，如下图 3-13 所示：

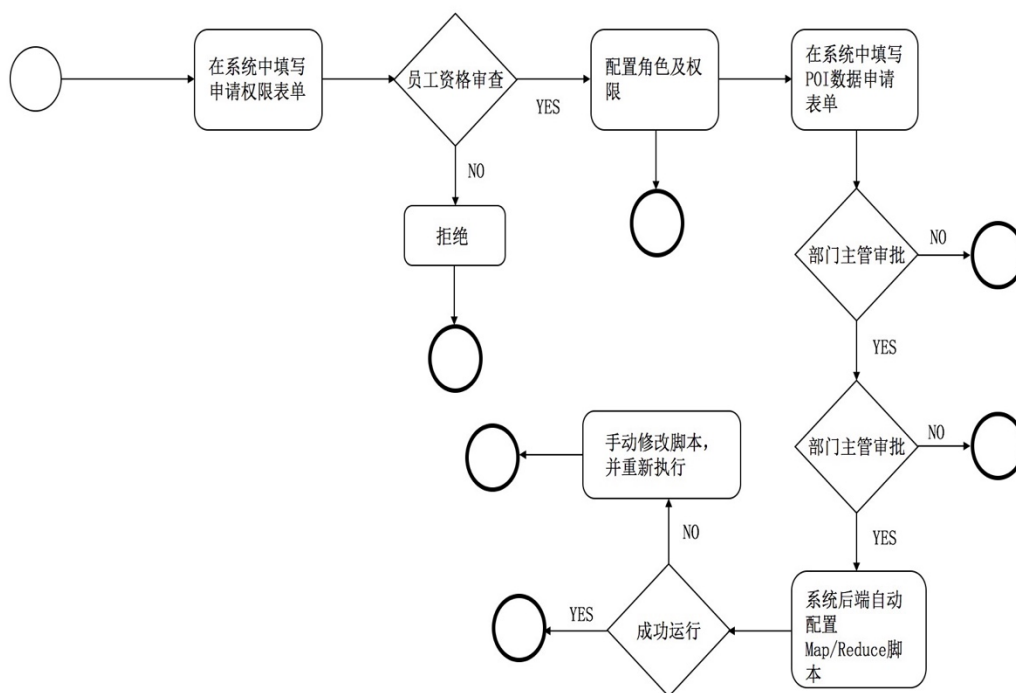


图 3-13 To-Be 模型总体业务流程图

3.5 系统需求分析

本章主要根据上面对现有业务流程的分析包括 As-Is 模型的建立，对该模型进行分拆成各个环节进行分析得出的优化及改进关键点。最后对各个环节提出解决的办法，并得出系统未来业务的 To-Be 模型。系统模型构建完毕，从而可以对本系统的实现做好相应的需求分析工作，下面将通过功能模块的划分，系统用例图的设计，主要模块的功能需求进行逐一介绍。

3.5.1 系统主要功能模块划分

根据前面章节的调研分析，本数据订阅系统的主要功能模块划分为：角色及权限管理、POI 数据申请管理、POI 数据审批管理、数据生产管理。如下图 3-14 所示：

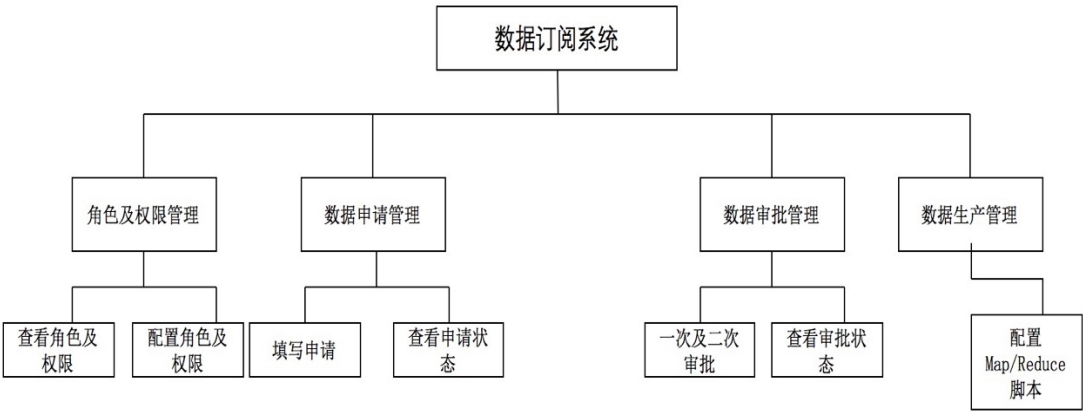


图 3-14 数据订阅系统主要功能模块划分

系统其他基础功能模块主要以服务形式存在，包括数据安全机制，系统即时通信机制。这些对于本系统的效率提升至关重要。

3.5.2 系统用例图

根据对系统的功能模块划分结果，得到整个平台的系统用例图如下，如下图 3-15 所示：

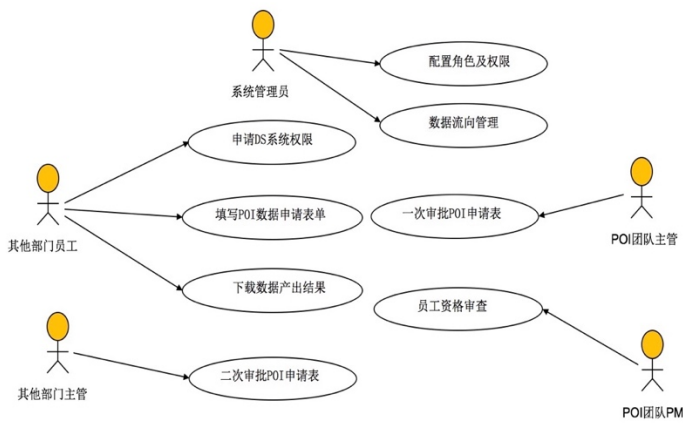


图 3-15 系统用例图

3.6 本章小结

本章节首先介绍了数据订阅系统本身，接着通过对现有业务流程建立 As-Is 模型并进行业务流程的利益相关者分析，得出系统的优化重点在于服务的提升一级办公自动化的实现。之后应用分析业务流程中的诸多科学的分析办法对本论文关注的业务流程进行细化分析，通过范围分析我们明确了项目应当完成的工作内容，确认实现的目标范围。通过收集各个角色的观点进行观点分析，了解系统优化的目标方向，根源分析方法则指出了根源问题的所在，瓶颈分析与情境分析则帮助找出业务流程优化的重点。基于 As-Is 模型进行分拆环节并根据之前分析的结果提出解决方法及改进途径，最后提出未来业务 To-Be 模型，为进一步的系统实现研究奠定了有力的基础。系统需求分析当中给出了系统主要模块的功能划分，以及整个系统的用例图。以上对整个业务流程的分析优化与设计，To-Be 模型在后续的系统实现当中也得到了充分的验证和肯定。

4 数据订阅系统整体设计与实现

本章节开始简单介绍 workflow 管理系统的体系架构以及一个参考模型的主要模块。主要基于上述对数据订阅系统业务现状的分析、业务的优化以及推出全新的设计思路，基于现有 workflow 管理系统的参考模型，实现过程定义，执行服务、管理及监控和客户应用的设计。最终完成数据订阅系统技术层面的整体设计与实现。

4.1 WFMC workflow管理系统体系架构

为了更好地支持业务过程建模, 分析和实施, 适应市场的需要, 一个完善的工作流管理系统体系架构能够从系统架构、术语的使用、接口实现多方面提供标准的定义。并以此为基础实现各个应用系统的集成。因此工作流管理委员会提出了关于工作流管理系统的规范。下图 4-1 给出了工作流管理系统的体系架构:

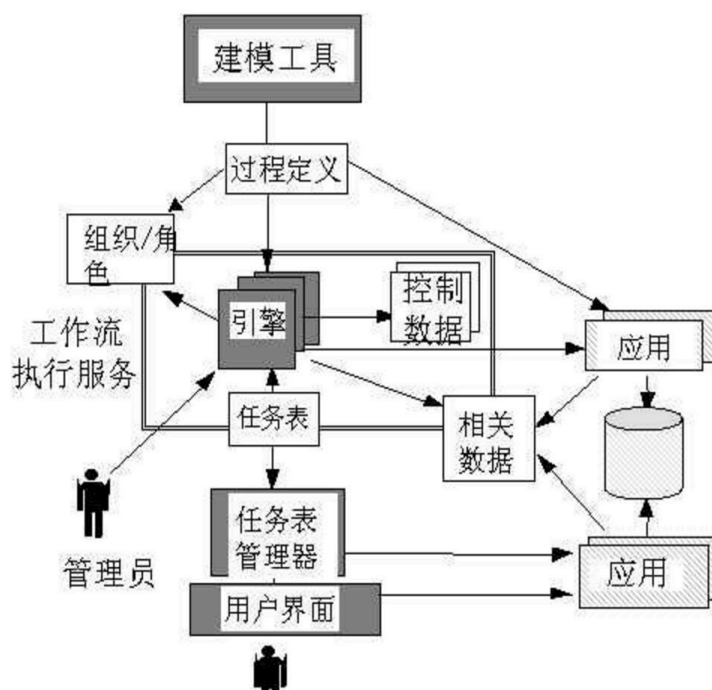


图 4-1 workflow管理系统体系架构

从上图 4-1 中, 可以知道, 一个完整的工作流管理系统主要有三类组件构成, 分别为:

- (1) 软件构件： workflow 管理系统中不同组成部分功能的实现

（2）系统控制数据： workflow 管理系统中软件构件使用的数据

（3）外部产品： 不属于 workflow 系统的组成部分，仅是被使用以完成整个系统的功能。

4.2 workflow 管理系统参考模型

workflow 管理系统参考模型给出了实现 workflow 技术的标准。在本参考模型中，详细描述了 workflow 管理系统的功能模块组成，即各个功能模块之间的接口功能定义。workflow 执行服务器负责 workflow 过程定义识别、实例创建、运行。过程定义工具是定义 workflow 模型；客户端应用程序表示 workflow 参与者查询自己的任务列表；管理和监控用来监督 workflow 的运行状态。workflow 管理系统参考模型见下图 4-2 所示：

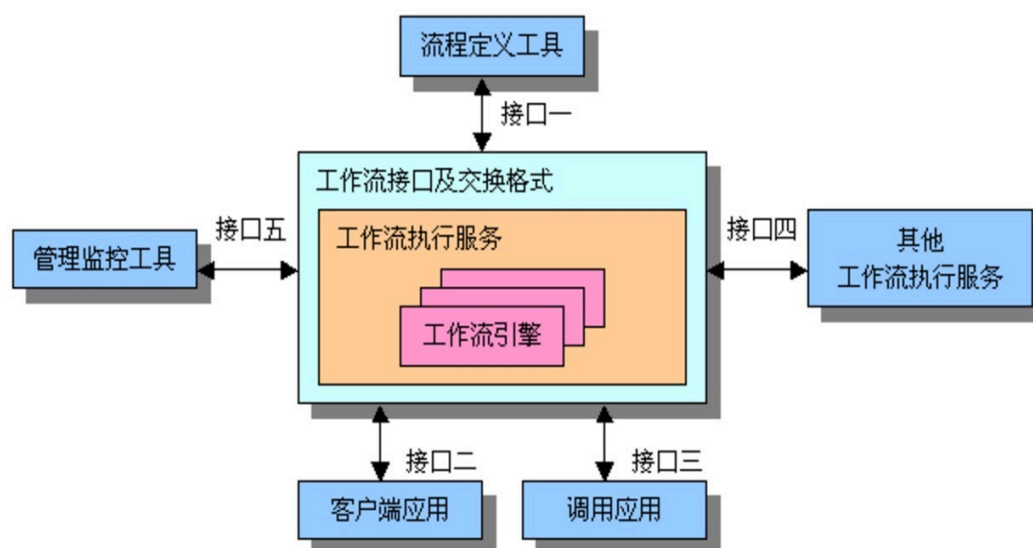


图 4-2 workflow 管理系统参考模型

4.3 数据订阅系统体系结构及系统组件组成

结合对本数据订阅系统的考虑，在参考 WFMC 的标准工作流管理系统的参考模型，考虑系统的实现复杂程度以及系统的易用性，笔者对本系统的结构进行重新设计，给出下图 4-3 适用于本业务的系统体系结构图：

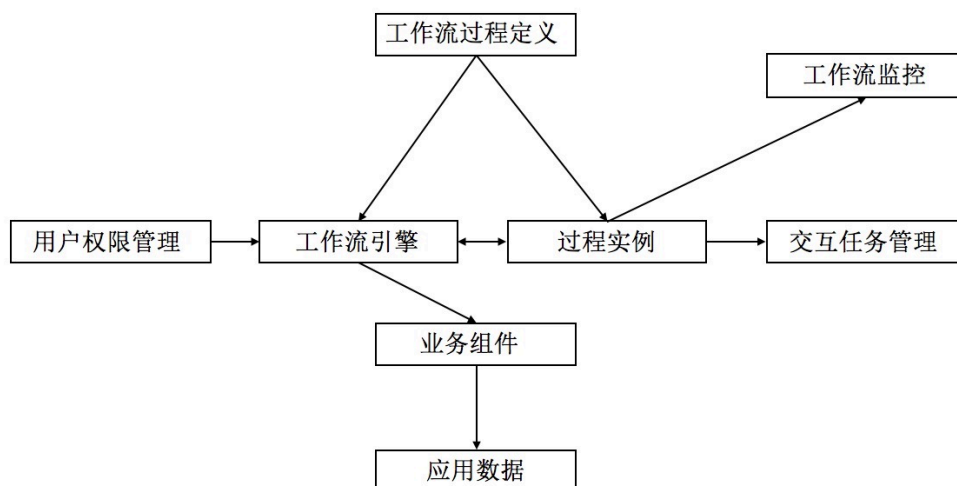


图 4-3 适用本业务的系统体系结构图

由数据订阅系统结构图可知，本系统主要组件分为以下三个部分：

（1）工作流过程定义：用于定制可视化页面的工作流过程模型（表现为事务处理过程）。

（2）工作流引擎：引导工作流自动化执行，作为系统守护进程级别运行服务。

（3）工作流监控：用于对当前运行以及已结束的流程过程的监控。

4.4 工作流管理系统各组件设计

基于上面给出的数据订阅系统的结构图，本系统的主要模块实现为三个部分：工作流过程定义（过程模型）、工作流引擎、工作流监控。下面将逐一介绍本系统的实现。

4.4.1 工作流过程定义

本系统的工作流过程模型能够形象化地表示出订阅流程的顺序及约束条件。该流程由一个无自环的有向图构成。组成元素包括有：过程、活动、控制连接弧、数据连接弧。

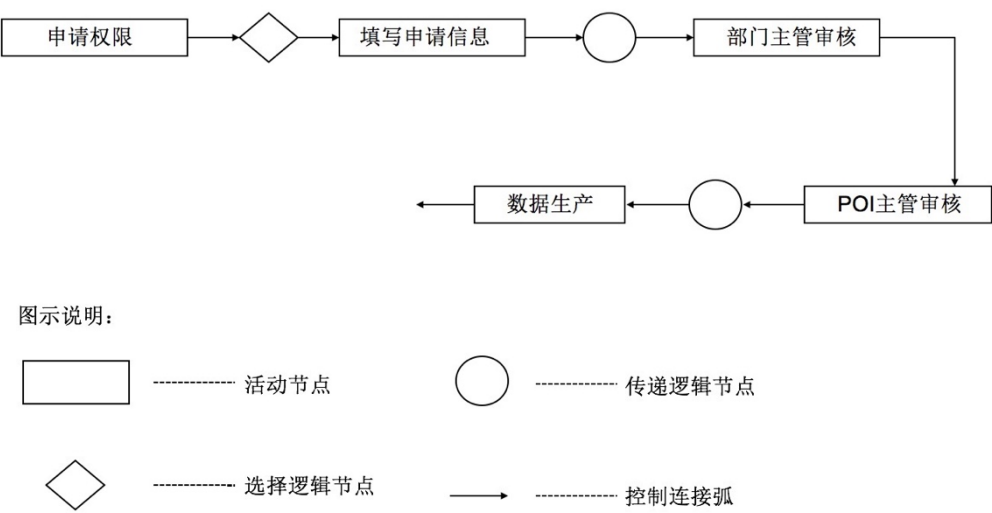


图 4-4 工作流过程模型

如图 4-4 所示的工作流过程模型，这里简单介绍下流程模型建立的过程：首先由部门员工向 POI 部门 PM 提出申请 POI 数据权限，这一步为选择型逻辑处理模式，任何一位 PM 都可以进行员工资格审查并予以处理。员工获得权限后，填写申请信息并传递给该部门主管进行审批，审批通过后进入 POI 主管审批任务列表项，POI 主管审批通过后进入数据生产环节，最后产出数据结果集并交付对方员工。

4.4.2 工作流引擎

工作流引擎作为运行服务组件中的核心层，中央控制中枢。实现本系统过程模型的识别，为工作流实例配置执行环境，包括各种状态：初始化状态、执行态、活动态、悬挂态、完成态以及终止态。按预定好的流程定义及数据信息引导实例的运行。

（1）初始化状态：过程实例已经生成，状态数据和工作流相关数据已经准备完毕，缺少促使过程执行的条件。

（2）执行态： 过程实例已经开始运行。

- (3) 活动态：一个或多个活动已经启动。
- (4) 悬挂态：过程实例处于静止状态并且没有活动激活，直到过程重新回到运行态。
- (5) 完成态：过程实例满足结束条件
- (6) 终止态：流程正常结束前，过程实例停止运行。

下图 4-5 给出过程实例的状态转换图：

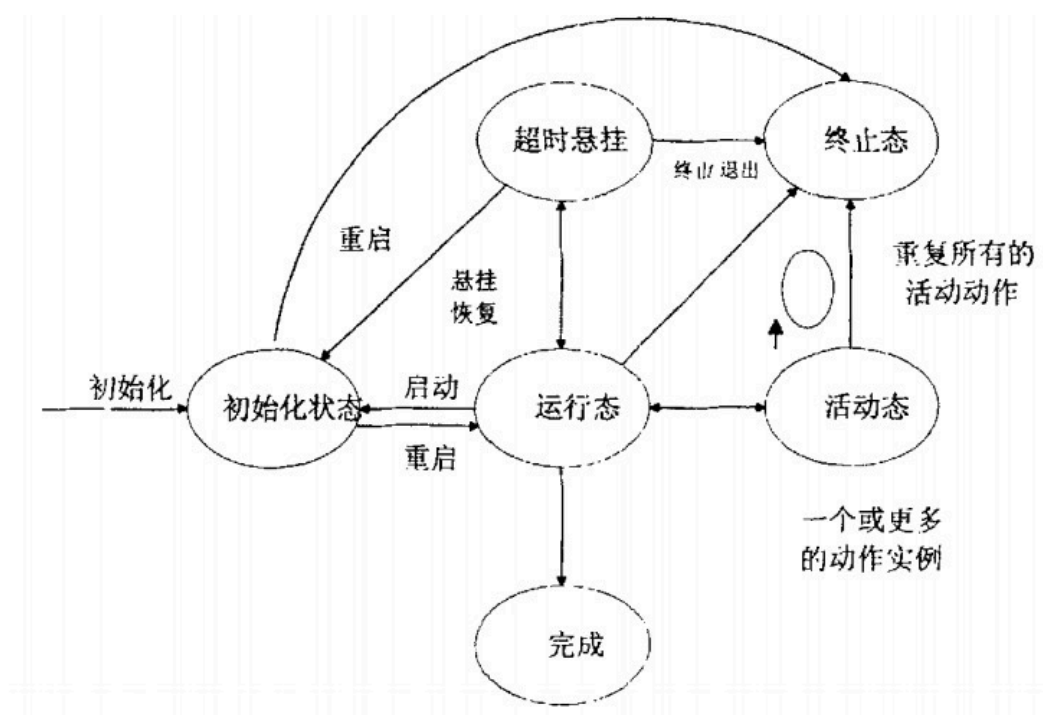


图 4-5 过程实例状态转换图

本系统的工作流引擎实现中，采用分层技术架构实现工作流引擎的开发，按模块分为下面三层：

第一层为用户接口层：主要为用户与工作流引擎间的交互模块功能实现。例如用户登录检测、工作任务列表查看、过程实例的创建维护及实例状态监控等。数据由底下提供，为工作流引擎提供的外部接口实现。

第二层为过程实例层：目的是给出过程实例的执行信息。例如执行状态及完成情况等。本层信息实时更新，且只涉及到该实例的相关运行数据，不包括有流程定义数据。过程实例推动过程中，主要通过调用底层过程定义逻辑模块功能实现。

第三层为静态模型层：为描述工作流过程定义，是 workflow 引擎内部自定义 workflow 模型描述方式。根据内部活动逻辑关系进行对实例的运行驱动，根据过程运行实时参数推动过程的执行。

4.4.3 工作流监控

工作流监控目的是帮助管理者对系统当前的过程实例运行进行实时跟踪监控。从而更加了解过程实例事务运行的情况，结果处理。采用 B/S 架构实现可视化界面监控。

该部分实现是通过从实例表中按实例号取出各个实例的基本状态信息并进行展示。定时器按时调用监控组件，得到实例表中各个实例的实时运行状态。下图 4-6 给出了基本过程：

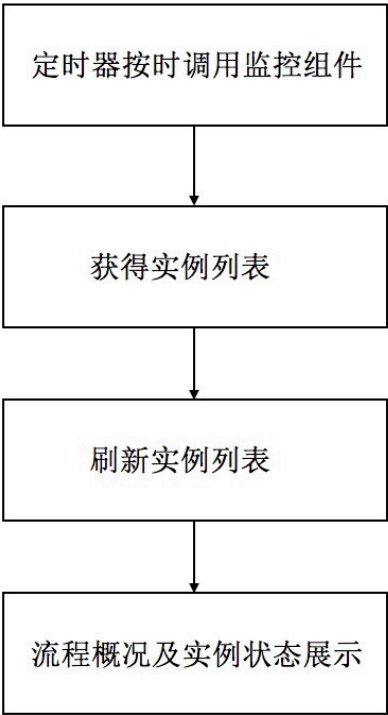


图 4-6 工作流监控基本流程

4.5 数据订阅系统角色图

数据订阅系统典型组织机构图如下图 4.1 所示。系统主要角色为：公司其他业务部门员工，公司其他业务部门主管，POI 团队主管，POI 团队 RD。如下图 4-1 所示：

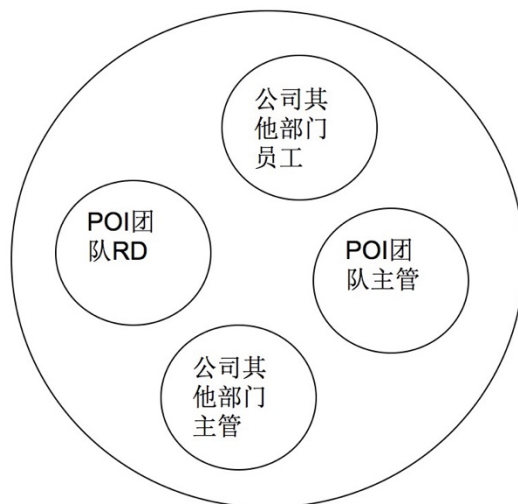


图 4-7 数据订阅系统角色图

系统的主要角色如上图框图表示，四大角色：公司其他业务部门员工，任务：在数据订阅系统上按业务需求如实填写申请表单，之后点击提交。公司其他业务部门主管，任务：在数据订阅系统上的审批管理区域对本部门内部的 POI 数据需求进行审批，下一站由 POI 团队主管进行审批。POI 团队主管，任务：对来自其他业务部门的数据申请单进行审批。POI 团队 RD，任务：负责维护和值班监管该数据订阅系统。

4.6 数据订阅系统主要模块功能划分

本节主要就数据订阅系统功能架构进行详细介绍。主要包括有系统功能模块划分图，主要分为角色及权限管理，申请管理，审批管理，数据安全的管理，数据生产管理四大部分。具体模块详见下图 4-8 所示：

4.6.1 系统整体功能框图介绍

根据数据订阅系统的整体规划，本系统的管理重点主要在于自动化实现系统的运行。主要分为下面几个部分：申请管理，流程状态管理，历史业务管理，审批管理，即时通信管理，角色及权限管理，流程监控管理，数据流向统计管理。数据订阅系统的主要功能模块包括如图 5-1 所示：

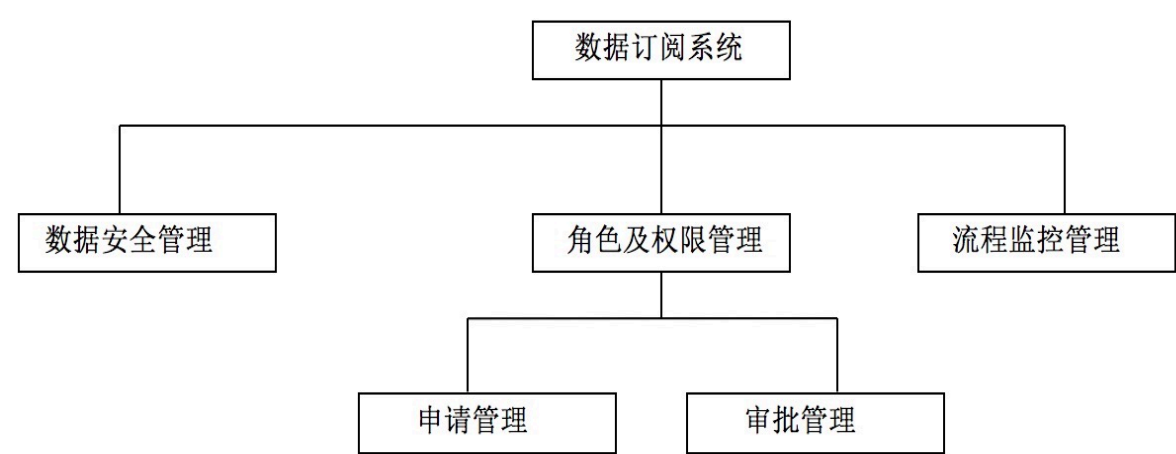


图 4-8 数据订阅系统的主要功能模块

4.7 系统表单模块实现

本节主要通过根据以上章节的分析，为实现系统的使用便利特性、在前端 UI 层面作了以下的设计，分功能模块：角色与权限管理，表单结构化，流程状态管理，审批管理。

4.7.1 角色权限管理模块

数据订阅系统的角色与权限管理较为简单，具体需要设三个角色，分别为：公司其他业务部门员工、公司其他业务部门主管、POI 团队主管。下面给出该模块的原型设计图，如下图 5-2 所示。



图 4-9 角色管理模块原型图设计

下面给出本系统的角色表结构：

表 4-1 数据库 Role 表结构

字段名	类型	NULL	Key	Default	Extra
role_id	int	NO	PRI	NOT NULL	auto_increment
role_name	varchar(20)	NO	FOREIGN	NOT NULL	
role_intro	text				

如上图 5-2 原型图所示，角色配置模块中需要根据账户来唯一确定一个用户，通过制定用户的角色，以及生效时间来完成配置用户的角色及其相应的权限。

4.7.2 数据申请表单模块

数据订阅系统的申请模块中，表单结构化管理是重中之重，公司其他业务部门员工需要通过该此表单进行按业务需求填写申请数据的元数据信息及必要内容。经过对上述章节的分析调研给出表单结构化的原型设计图，如下图 4-10 所示：

数据订阅系统

https://www.example.com

项目名称

xxxxxxx

用途说明

上级经理

xxx

数据来源

选择主来源

选择子来源

输出字段

xxx

xxx

xxx

筛选条件

选择字段

选择符号

此处填条件值

条件关系

and

添加筛选条件

取消

确定

图 4-10 数据订阅系统表单结构化原型设计图

如图 4-10 所示，数据订阅系统表单结构化原型设计图主要包括有：项目用途说明，上级经理账号，数据来源（分别有主来源和子来源），输出字段的选择，增加筛选条件。表单信息主要存储在业务数据库中。

4.7.3 定时器模块

公司其他业务部门在数据订阅系统申请数据时，一般期望一次申请，可以指定多次分时间推送新数据。因此定时推送方案中的 UI 方案设计显得尤为重要。定时推送方案设计中分为三种周期进行推送：天级别，周级别，月级别。根据表单提交的信息配置 Crontab 执行命令。下面将给出三个级别的 UI 原型设计图。

1) 天级别

天级别的周期设定，只需给出时间设定一览，帮助用户可以指定每天的某个时间点开始生产数据并完成推送。Crontab 命令为：* 12 * * * myCommand.



图 4-11 天级别定时 UI 设计图

2) 周级别

周级别的周期设定，需要给出星期选择以及时间设定一览，帮助用户可指定一周当中的某一天的某个时间点开始生产数据并完成推送。

Crontab 命令为：* 12 * * 1 myCommand.

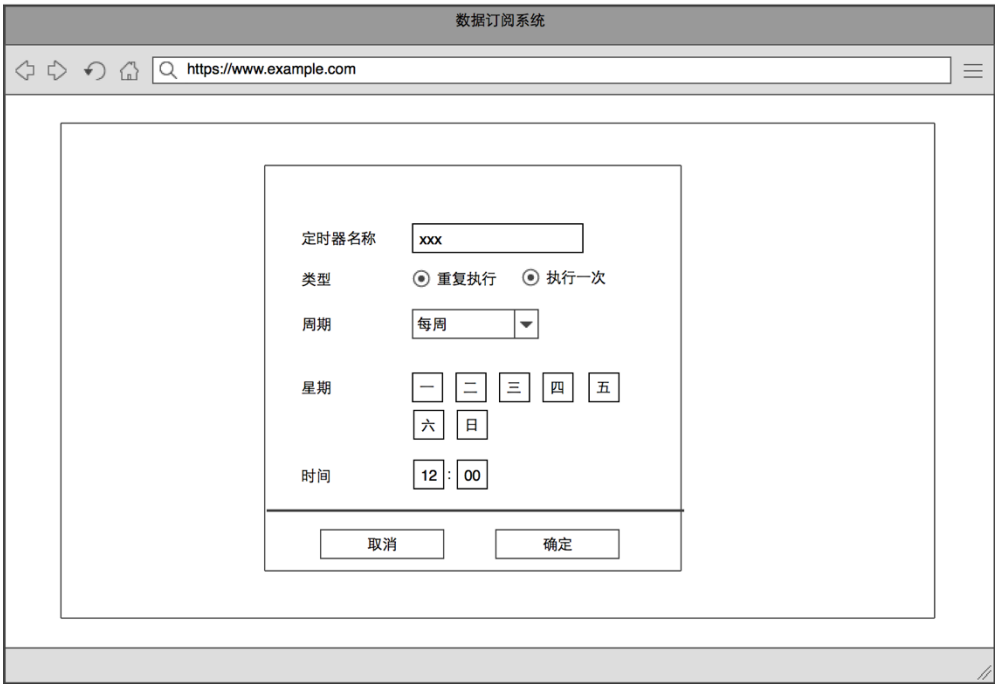


图 4-12 周级别定时 UI 设计图

3) 月级别

月级别的周期设定，需要给出日期选择以及时间设定一览，帮助用户可以指定某个 月中的某一天的某一个时间点开始生产数据并完成推送。
Crontab 命令为：`* * 1 * * myCommand.`

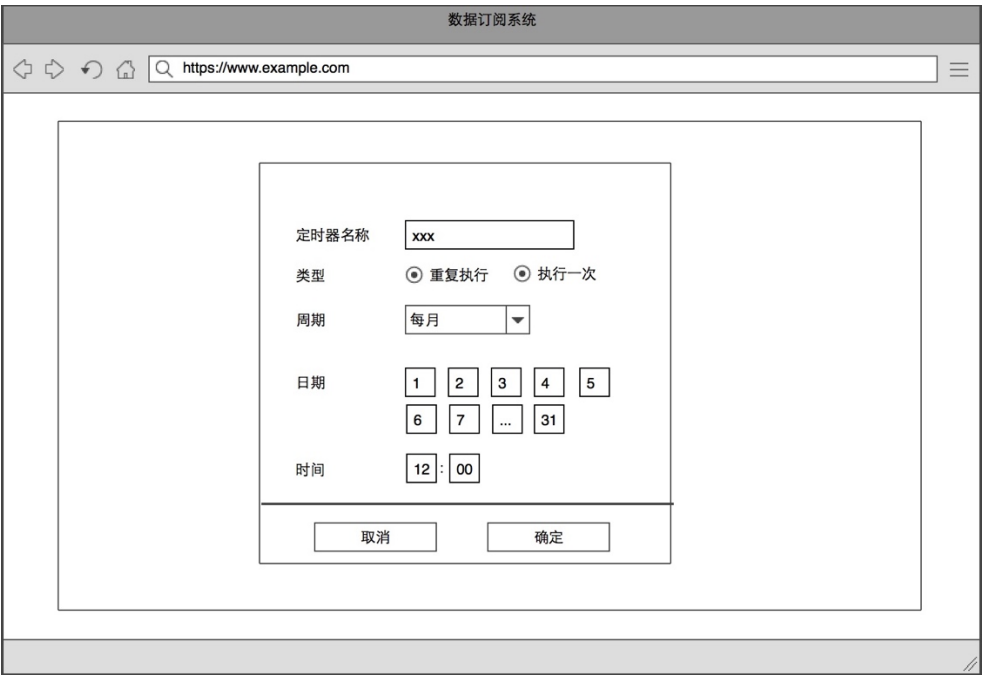


图 4-13 月级别定时 UI 设计图

三种级别的定时器方案帮助公司其他业务部门的员工能够自由定制数据申请元数据的需求及条件，实现业务流程自动化运行，提高业务执行效率。

4.8 WebSocket 即时通信设计与实现

本小节主要介绍业务流程当中业务沟通的消息通讯方案的构造过程，根据先前分析可知传统的电子邮件传输通信的方案给业务执行带来极大的不利，原因在于邮件内信息内容多而杂，与实际业务流程呈分散状态，且存在传输安全问题。因此通过调查 WebSocket 全双工通信可以达到即时通信的目标，根据相关资料的检索发现，WebSocket 存在低版本兼容问题，在大量检索资料的同时，发现一个很有效的 Socket.io 库可以解决因浏览器兼容问题。下表 4-2 主要给出现有浏览器对 WebSocket 协议的支持情况：

表 4-2 浏览器支持情况

浏览器	支持情况
Chrome	Supported in version 4+
Firefox	Supported in version 4+
Internet Explorer	Supported in version 10+
Opera	Supported in version 10+
Safari	Supported in version 5+

由于 WebSocket 在多种浏览器中存在兼容性问题，目前的 Socket.IO 库解决了多种浏览器兼容问题，当遇到无法支持 WebSocket 时，会进行优雅降级处理。Socket.IO 是基于 JavaScript 实现的支持 WebSocket 的常用在即时通信服务系的研发开源工具。Socket.IO 还封装了轮询方法以及其他的即时通信方式。并且在服务端完成即时通信代码的开发工作。Socket.IO 实现的 Polling 通信功能也包括 Adobe Flash Socket、AJAX 长轮询、AJAX multipart streaming、持久 Iframe 以及 JSONP 轮询等。

Socket.IO 已经具有众多的强大的功能模块以及扩展 API，诸如，`session.socket.io`、`socket.io-cookie`、`session-web-sockets`、`socket-logger`、`websocket.MQ`、`socket.io-mongo`、`socket.io-redis`、`socket.io-parser` 等。其中 `session.socket.io` 是 http session 中间件，用于 session 相关操作，`session-web-sockets` 是以安全的方式进行传递 session，`socket.io-parser` 为客户端与服务端之间通讯的默认协议的实现模块^[11]。

数据订阅系统的即时通信场景有两种，下面通过演示公司其他员工在填写并提交申请表单时运用 WebSocket 即时通信的案例，分两种情况，其一是 POI 团队主管在线时，表单提交消息实时传递给 POI 团队主管线上并且也记录到数据库 Message 表。下图 4-7 为在线时通信设计。

4.8.1 数据库 Message 表设计

数据库 Message 表的用途是作为存储系统之间通讯消息的目的，包括提交申请表单的消息通知，一级审批通过的消息通知以及业务执行结束的消息通知等。当对方用户在

线时通过 WebSocket 全双工通信的方式将消息即时发送过去，并写入数据库表 Message 表，将读取状态字段值置为未读取。当对方用户读取该消息时，即将数据库表中的读取状态字段值改为已读取状态。下表 4-3 及 4-4 将分别给出 User 表及 Message 表的表结构设计。

表 4-3 数据库 User 表结构

字段名	类型	NULL	Key	Default	Extra
user_id	int	NO	PRI	NOT NULL	auto_increment
department	varchar(40)	NO		NOT NULL	
user_name	varchar(40)	NO		NOT NULL	
role_id	int	NO	FOREIGN		
acc_status	int			1	

User 表主要记录用户的基本信息，由于登录验证已经由公司的统一账号平台提供服务，本系统只提供对用户权限的管理，通过 User_id 唯一识别一个用户。Department 为科系名称，与 user_name 可以唯一标识一个用户。Acc_status 字段为标识用户合法状态，为 0 则为禁用，1 为正常使用状态。

表 4-4 数据库 Message 表结构

字段名	类型	NULL	Key	Default	Extra
id	int	NO	PRI	NOT NULL	auto_increment
apply_id	int	NO	FOREIGN	NOT NULL	

submit_time	date	N0
-------------	------	----

status	int	0
--------	-----	---

apply_id 是外键引用数据库中 apply 表的主键，submit_time 为提交时间，status 为读取状态，默认为 0 表示未读取，1 为已读取。

4.8.2 WebSocket 即时通信业务类图设计

该通信模块中业务基本流程为：用户提交申请表单，将表单元数据及条件信息插入数据库 apply 表中，并且与 WebSocket 通信接口建立连接检测对方用户是否存在，若存在则应用在线通信设计方案，否则应用离线通信设计方案。下图 4-14 为笔者给出的类图：

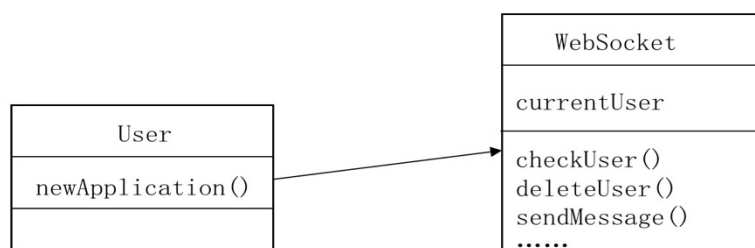


图 4-14 WebSocket 通信类图设计

4.8.3 在线通信设计方案

当 POI 团队主管为在线状态时，申请数据表单提交时，在选择项目经理一项时，系统前端通过 ajax 异步请求后台数据库用户表中对应姓名及部门的经理信息，包括有 user_id, user_name, department 信息。在提交表单后，系统对申请数据元数据及条件信息插入数据库中，插入成功后，检测 socket session 中的 socket 名称为该 user_id 是否存在，存在即标识在线。系统则将新表单消息写入 socket 并发往对方。并且也需要将该条信息写入到系统数据库提前设计好的 Message 表，标记读取状态为未读取。

● POI团队主管在线

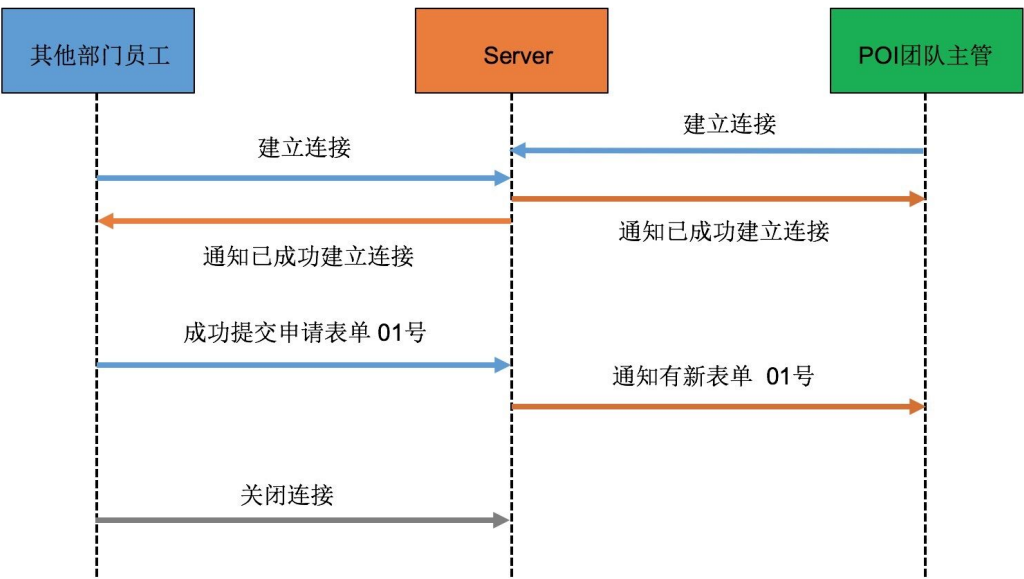


图 4-14 POI 团队主管在线时系统通信设计

4.8.4 离线通信设计方案

当POI 团队主管为离线状态时,系统无法进行实时 socket 读写发送,因为当前 server 与 POI 团队主管并未建立 WebSocket 连接,因此需要将该条信息写入到系统数据库提前设计好的 Message 表, 标记读取状态为未读取。

● POI团队主管离线

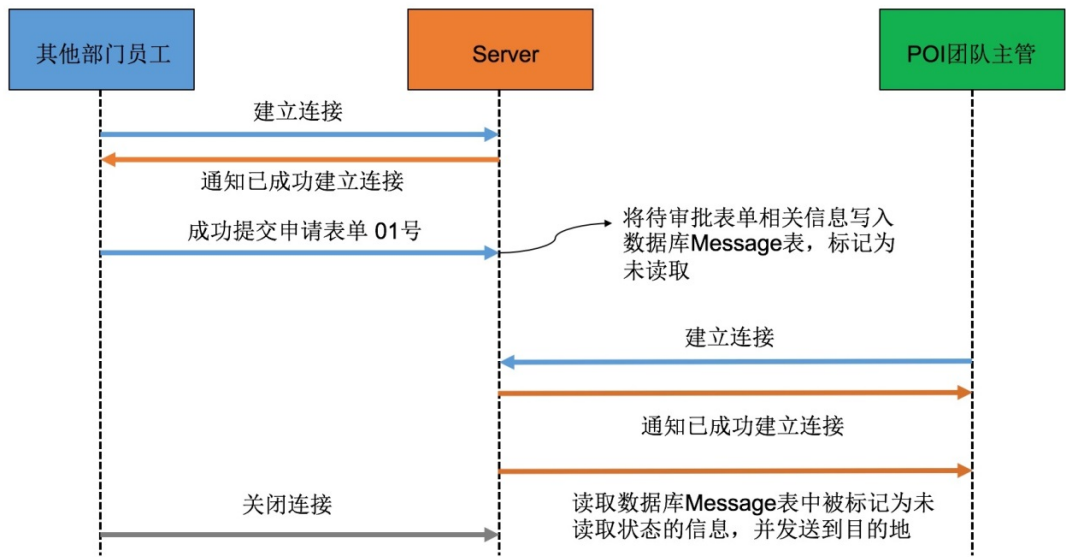


图 4-15 POI 主管离线时系统通信设计

如上面所给出的两种针对在线及离线通信设计方案中运用了 Socket. IO 库，有效地解决了低版本浏览器兼容问题以及在线，离线两种情况下的稳定通信保证。

4. 8. 5 WebSocket 即时通信业务时序图

下图 4-16 为笔者针对前面对即时通信的两种方案设计（在线及离线），提出的对于系统代码相关联的通信业务时序图，其中技术方案前提是维护一个WebSocket 在线队列，且 socket 名称为各个用户在系统数据库中 User 表的 user_id，唯一识别一个用户。详见下午 4-16 所示：

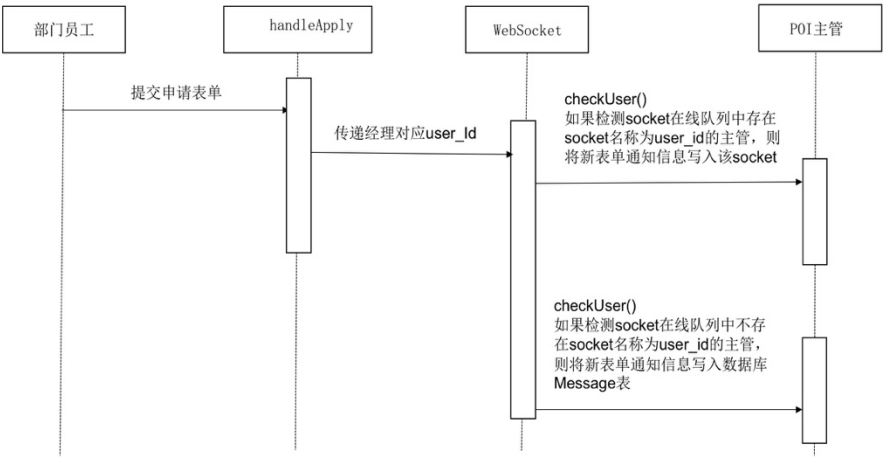


图 4-16 WebSocket 通信业务时序图

4. 9 本章小结

本章开头介绍了本系统应用的工作流技术，介绍了 WFMC 的标准 workflow 管理系统体系架构以及 workflow 管理系统的参考模型并根据业务本身系统的需要提炼出数据订阅系统的体系结构及针对 workflow 管理系统的三大组件进行开发，流程定义、workflow 引擎以及 workflow 监控。后面详细介绍了基于前面几个章节的业务流程分析结果以及最初优选的技术途径开始整个暑假订阅系统的研发及架构设计与实现工作，首先介绍了本系统的主要角色有，公司其他业务部门员工、公司其他业务部门主管、POI 团队主管、POI 团队 RD。通过功能树介绍了本系统研发的目标功能，主要有角色及权限管理、流程状态管理、表

单系统、定时器、系统消息即时通讯模块。重点介绍了定时器分三种周期级别设计，天、周、月。帮助其他业务部门员工方便获得最新数据，而无需劳力重复之前申请操作。最后也重点介绍了为解决系统即时通信给出的两种方案，针对在线及离线情况并给出类图以及时序图的设计，解决思路是通过维护一个 WebSocket 在线队列，并且 socket 名称为唯一的存在于数据库 User 表中的 user_id。

5 系统分析及优越性验证

本章节主要是基于之间章节对现有业务流程的分析，对数据订阅系统进行整体分析并给出优选技术方案的优越性验证。通过利用量化手段来进行验证。同时会对系统的风险及时间实施之后的管理提升进行了阐述。

5.1 系统业务流程优化 KPI 分析

本小节主要给出作者本人对系统的 KPI 评估和分析，重点从 KPI 整体目标及各个方面展开进行提炼和描述。数据订阅系统的 KPI 目标不仅仅是要确保系统的正常运转，还要同时把控好投入的各种成本，监控系统项目的实时进度安排，需要保证项目研发的时效性，尽最大能力减少风险。

系统项目 KPI 测试的方法主要如下：

项目范围：把握业务流程方向以及需求的变更。

项目进度：关注项目里程碑的突破和实现。

项目质量：管理和维护好质量报告单。

项目风险：通过多种策略，维护系统风险报告单。

人为因素：努力将人为因素导致的失误降低到 10 以下。

业务效率：提高业务执行效率，实现高度自动化，时间缩短到 60%。

5.2 系统业务升级的管理提升

一旦当前面几个章节所涉及到的短期及长期目标都顺利实现之后，必然会给整个团队的数据管理带来非常大的提升和改变。主要提升改变有以下几点：

（1）系统业务流程当中的申请单内容由原先的非结构化电子邮件内容组织方式转变为系统数据库结构化的信息存储，有利于系统对申请管理业务查询，数据流向管理。

（2）业务流程都被实时监控，申请者及管理者都可以更好地实时了解业务的运行状况，帮助后续工作安排等，提高工作效率。

（3）各个业务部门的主管以及 POI 团队的主管可以更好地管理本部门员工的数据申请任务执行情况，可以更好地做好评估和审查工作。

（4）数据订阅系统的新上线使得原有业务流程焕然一新，更方便，也帮助 POI 团

队数据信息化管理朝前迈了一大步。

5.3 业务执行效率的提升

本节通过利用量化手段对新系统上线后业务执行效率进行评估验证。给出了业务执行过程中各个角色的用时统计。以及给公司其他业务部门的项目带来的有效收益包括业务项目研发效率。详细查看下表 5-1 业务执行效率提升评估表：

表 5-1 观点分析概括表

系统角色	新系统前后效率评估
公司其他业务部门员工	<div>1. 原先填写一个符合要求的申请邮件以及来往沟通的时间需要耗费 15 分钟左右。</div> <div>2. 预估新系统上线后, 在理清申请数据元数据信息后, 填写申请单仅需要花费 5 分钟左右时间。</div>
公司其他业务部门主管	<div>1. 在审批部门内的数据申请邮件过程中发现, 因为每天收发邮件数量大, 常常容易错过。</div> <div>2. 新系统对该业务流程中的消息通讯进行统一管理, 方便审批快速操作。</div>
POI 团队主管	<div>1. 通过电子邮件方式进行抄送审批的方式, 受制于电子邮件本身固有特点: 大量混杂的电子邮件, 无法及时有效得到通知并加以处理。</div> <div>2. 采用即时通信的方式, 系统管理业务内部消息通讯, 帮助审批人员更加快速及时完成审批工作。</div>
POI 团队 RD	<div>1. 原先需要手动调用脚本传参执行脚本拉取数据,</div>

大概整理申请元数据信息等整个过程需要花费 15 分钟左右时间。

2. 新系统上线后实现自动化配置脚本去拉取数据的话，不需要 RD 再耗费任何时间或者精力去跟进数据的申请工作，只需要定期维护和管理，保证系统正常运行即可。

5.4 系统优化实现的风险评估

风险评估的目标在于标识，定位以及消除各种风险因素带来灾难性后果，在其发生之前可以进行有效阻止。根据以上的风险内容定义，作者对系统的风险进行了整体的评估。下面将列举出系统项目进行当中所有可能遇到的风险。下表 5-2 主要描述针对每一种可能的风险，给出相对应的减少风险的解决措施。

表 5-2 风险预测及解决措施表

可能遇到的风险	解决措施
系统解决方案不符合预期	1. 详细的项目规划，强大的项目管理，认真审查及良好的沟通机制
系统实施之后业务的连续性	1. 保证之前的业务流程不被当前新业务流程干扰，作为灾备系统使用。

5.4 本章小结

本章节为验证项目优越性，提出之前的 KPI 的预期目标。通过利用量化手段来进行验证。本章第三小节中给出了对于系统主要角色在新旧系统上业务执行效率的评估提升，同时对也系统的风险及时间实施之后的管理提升进行了简单的介绍。

6 结论

作者通过对原有数据订阅的业务流程进行各项科学的分析,得到现有业务 As-Is 模型,并找出原有业务流程存在的诸多关键问题,主要归纳为以下几点: 1. 业务沟通效率低、沟通方式不灵活 2. 系统大部分工作为手动操作,人为因素过多,业务执行效率低 3. 数据信息化管理存在申请管理漏洞等。接着通过应用结构化 workflow 系统的管理,进行流程定义、workflow 引擎开发设计、workflow 监控设计,进行优化和自动化实现。提出解决问题的方法和改进点,提出未来业务 To-Be 模型。论文中第四章给出了系统重要模块的设计方案实现内容,在第五章给出了新系统上线的优越性验证。基本通过 KPI 预期的目标,由于项目时间以及人力有限,系统的进一步优化为长期目标,将会在之后不断升级过程中实现。本人希望自己所做的关于数据订阅的业务分析与研究能够为今后类似订阅业务流程的优化与自动化实现相关的研究提供一种思考方法。

参考文献

- [1] 吴小丽. 浅谈企业信息化[J]. 科技情况开发与经济, 2007, (7): 86-88.
- [2] 邓启军. 基于 Web 和工作流的企业自动化管理系统, 2010, (5): 1-2.
- [3] 范玉顺. 工作流管理技术基础[M], 北京:清华大学出版社, 2001
- [4] 田光明. 情景分析法[J]. 晋图学刊, 2008 (3): 98-102.
- [5] 王春雨. 融资租赁业务系统优化改进方案的研究与实现[D]. 北京交通大学, 2012
- [6] 胡. 基于工作流的文档管理系统的研究与开发. 硕士学位论文, 2002
- [7] 木桶原理[EB/OL]. (2009-5) [2016-5]. <http://wiki.mbalib.com/>
- [8] 基于 AWS 技术实现发布/订阅服务[EB/OL]. (2013-8) [2016-5]. <http://www.infoq.com>
- [9] Cron 技术[EB/OL]. (2015-9) [2016-5]. <https://zh.wikipedia.org/wiki/cron>
- [10] Crontab 定时任务[EB/OL]. (2015-3) [2016-5]. <http://linuxtools-rst.readthedocs.io>
- [11] Socket. IO[EB/OL]. (2014-10) [2016-5]. <https://zh.wikipedia.org/wiki/Socket.IO>
- [12] 企业工程[EB/OL]. (2015-12) [2016-5]. <https://zh.wikipedia.org>
- [13] Hypertext Transfer Protocol[EB/OL]. (2016-5) [2016-5]. <https://en.wikipedia.org>
- [14] WebSocket[EB/OL]. (2016-4) [2016-5]. <https://en.wikipedia.org>
- [15] Real-Time Publish/Subscribe[EB/OL]. (2005-1) [2016-5]. <https://www.rti.com>
- [16] 林郎碟. 实时发布订阅系统及其相关技术的研究[J]. 福建电脑, 2010, (1): 5-8
- [17] 包心如. 面向消息的实时发布订阅系统的设计与实现[D]. 北京邮电大学, 2011
- [18] 赵伟. 发布/订阅系统数据模型的设计与实现[D]. 南京理工大学, 2009
- [19] 刘健. 航天大数据平台数据订阅系统设计与实现[D]. 西安电子科技大学, 2015
- [20] 冀素琴. 基于工作流技术的办公自动化研究[J]. 中国优秀博硕士论文, 2001

致 谢

感恩，本论文最终能够顺利完成，离不开作者导师 xxx 的细心指导和帮助。在论文最初的选题时，xxx 在毕设小组会议上为我们梳理各自选题的思路，指导我们以科学的方法在实习工作当中提炼有意义的论文题目。在完成毕业设计过程中，每一次的研究陷入瓶颈时，老师的指导都给我非常大的帮助，使我可以重新恢复思考并且最终得出成熟的方案设计。导师渊博的学识，认真严谨的治学态度也使作者获益匪浅。值此论文交付之际，在这里向我敬爱的导师致以崇高的敬意和由衷的感谢！

特别感谢单位给了我研究和实践的机会，在项目的实际开发过程，公司的导师，实习小组的同事也给了我许多的帮助和指导。感谢导师给我机会去主导本系统的架构设计，让我更加深入的了解原业务流程的问题以及痛点。感谢同事们的配合和支持，使我能够顺利地完成任务，在此，向帮助过我的每个人表示由衷的感谢！

最后也特别感谢软件学院和各位老师在这四年当中对我的培养。在专业知识、理论基础、工程思想以及综合素质上各方面的提升离不开学院老师的辛勤付出。感谢四年来陪伴在作者身边一起学习，探讨技术，共同成长的同学和朋友们。

感谢评阅作者毕业设计论文的各位老师，感谢你们在百忙中抽出宝贵的时间认真严谨地对论文进行评阅！

附 录

附录 A 原文及翻译

Workflow Management Systems: Formal Foundation, Conceptual Design, Implementation Aspects

Chapter 1 Introduction

Workflow management aims at modeling and controlling the execution of application processes in dynamic and heterogeneous organizational and technical environments [67, 32]. Hence, the motivation for workflow technology comes from the specific requirements of information systems applications. To provide a concise motivation for workflow management in general and for the topics addressed in this thesis in particular, specific properties of information systems applications are introduced. Based on these properties, important research issues in workflow management are derived, and a formal foundation of workflow management is introduced, which serves as a basis for solving these issues. A key aspect in computer science is developing theoretical results and showing their validity in operational prototypical software systems. Based on theoretical results with respect to the research issues addressed, this thesis discusses the conceptual design of workflow management systems. The concepts and methods introduced are validated in prototypical implementations of workflow management systems.

Background, Motivation, and Research Issues

The last decade of the 20th century has brought severe changes to how organizations in commerce, public administration, education, and in science and engineering operate, how they interact with other organizations and with their clients. Many of these new developments are fueled by recent technological advances, for example the availability at low cost of computing and network resources and, most prominently, the rapid growth of the Internet. As a result, individuals can now communicate conveniently with colleagues and friends in other parts of the world, characterized by the global village metaphor. In the business sector, technological advances have created what is now seen as the global economy. 5 6 Chapter 1: Introduction Probably the most obvious appearance of these developments is the great interest in online applications, i.e., attracting potential clients on the Internet, accepting orders, exchanging product specifications or scientific data and, to some extent, providing online delivery of goods and facilitating negotiation and payment via the Internet. In these settings, information, products, and services are now available to clients world wide instantaneously after they are released. This property of online applications have strong implications on their design, implementation, and maintenance of information systems supporting these applications. In particular, it motivates the need for fast creation of new services and the flexible adaptation of existing ones. On a

technical level, system failures are no longer acceptable since they are immediately exposed to clients, who are likely to choose another organization to provide the desired services. As will be motivated in the remainder of this chapter, these requirements impose interesting research issues in workflow management, which can be solved by using a variety of computer science concepts, methods, and techniques. Before discussing the research issues addressed in this thesis and the concepts, methods, and techniques used for solving them, some general remarks on the application background of workflow management systems are appropriate.

From an application-oriented point of view, an important approach to cope with the new requirements is the explicit modeling of different aspects of applications and of the environment in which they are executed. The most prominent individual aspect in this context is process orientation, which has become a widely accepted and a successful approach to describe how work is performed in organizations. In the process orientation approach, application processes are in the center of attention rather than individual functions performed during these processes. To motivate process orientation, a brief look at the traditional approach to model work in organizations is taken. In the traditional approach, work is decomposed into fine granules, called functions. Work is organized around functions, and the relationship between functions is typically not taken into account. Despite of the fact that functions are performed properly by the individual applications, work in these settings is often performed inefficiently. Reasons for this fact include a variety of media to store application-specific data as well as redundant work and waiting times introduced by poor management of the interrelationships between individual functions.

As was realized in manufacturing earlier where production lines have been introduced decades ago to enhance the productivity of manufacturing goods, work is typically performed in processes. This means that the relationships between individual activities have to be taken into account, and work should be organized around processes rather than around individual functions. This observation also holds for processes which manage information rather than physical goods; these processes are called information processes. 1
Introduction 7 Information processes appear in the domains mentioned above. In commercial organizations, insurance claim management and credit request processing are typical examples of application processes. A sample application process in public administrations is processing a proposal to construct a private house. Student enrollment in a university involves a number of steps, which can be characterized as an application process in the education sector. Laboratory information management, experiment management, and the construction of complex apparatus are typical examples of application processes in science and engineering.

The explicit modeling of processes first appeared in them context of applications in commercial organizations; application processes in these settings are called business processes. The notion of a business process is defined by Hammer and Champy [42] as “a collection of activities that takes one or more kinds of input and creates an output that is of value to the customer”. Hence, business processes typically involve mission-critical processes of an organization. A good indication of mission-criticality is that the organization was founded to perform these processes [68]. The activity of modeling business processes is called business process modeling. It is worth noting that business process modeling is not restricted to application processes in commercial organizations. In contrast, it can also be applied to other areas in which application processes need to be modeled, for instance in public administrations, education, and in science and engineering. Abstracting from the particular domain, the terms application process and business process can be used. The design of business process models is typically developed by business domain experts rather than by computer scientists. Issues related to business process modeling will therefore not be addressed in this thesis.

After an organization has captured its business processes, software development projects are faced with the decision on how to support these processes properly. At this point workflow technology comes into the picture. As indicated above, workflow management aims at controlling the execution of the automated parts of business processes. Hence, it has to offer formalisms to express the automated parts of business processes and the technical and organizational environment in which they are executed. In general, these formalisms are known as workflow languages. Using workflow languages, workflow specifications are created, which are used as inputs for workflow management systems. Therefore, workflow languages have to be formal, so that the workflow management system can unambiguously control workflows according to given workflow specifications.

Based on the specific requirements of information systems applications, an analysis of the conceptual work as reported in the literature, and the limitations of existing workflow management systems, research areas in workflow management and corresponding functionality of workflow management systems are described: 8

Flexibility and Dynamic Adaptations: While flexibility has been one of the major motivations for workflow technology from the beginning, only limited functionality with respect to flexible workflow management is provided by traditional workflow management systems. Today, the need for additional flexibility is widely accepted in the workflow community, indicated by a number of recent workshops [101, 61, 19].

Integration, Re-Use, and Software Components: The re-use of workflow specifications to assemble new workflow specifications from existing ones is an important aspect in workflow management. A formal foundation of workflow specifications is an important prerequisite in this context. On a more technical level, the integration of external applications and existing software components is a promising approach.

Scalability, Availability, and Distributed Workflows: Scalability and availability are important aspects which have to be satisfied by information systems, indicating the ability to enhance the throughput by installing additional resources, and providing services without failures, respectively. Distributed workflow management is a promising approach to enhancing scalability and availability of workflow management systems, as will be discussed in detail below.

In the remainder of this section, these research areas are motivated in some detail, research issues are identified, and conceptual and technical requirements of suitable workflow support are derived. **Flexibility and Dynamic Adaptations** Given the dynamic structure of today's organizations in commerce, public administration, education, and in science and engineering, it is unlikely that application processes are modeled once to be executed repeatedly without any changes. On the contrary, processes may change gradually over time to reflect changes in the environment of the process — processes evolve. Of course, flexibility is not restricted to software systems. According to the dictionary, an artifact is flexible if it can "easily be changed to suit new conditions". Flexibility in workflow management is used with this meaning: A workflow application is flexible if it can easily be changed to suit new conditions, imposed by changes in the environment of the application process. Depending on the application domain under consideration, there may be quickly evolving application processes and there may be slowly evolving ones. In some application domains there may even be processes which do not change at all, for instance where process structures are

defined by legal regulations, which 1 Introduction 9 are not expected to change over long periods of time. For these application processes traditional workflow support will suffice, for others, it won't.

As indicated above, flexibility has been a major motivation of workflow management from the beginning: By explicitly modeling processes and a variety of additional aspects to be discussed later in this thesis, individual dimensions in workflow specifications can be subject to change without changing other aspects. As will be indicated below, work-flow languages are designed in a modular way that allows to perform changes to certain aspects, while preserving other aspects.

In general, workflow management systems provide the flexibility to perform changes to workflow specifications, and all future workflows will use the new, improved work-flow specification. However, workflow specifications for given workflows cannot be changed in traditional workflow management systems. This means that once a work-flow has started, no changes can be applied to its structure. This is a severe limitation of workflow management systems, since workflows can be long-running activities, which may run for days, weeks, or even for months or years. Restricting modifications to new workflows would render workflow technology virtually unusable in these settings, since long-running workflows would have to continue with a superseded workflow specification for long periods of time. It would be much more adequate to allow running workflows to be adapted to new and improved workflow specifications. Hence, the requirement to react quickly to changes in the environment of a process is not matched by the functionality of traditional workflow management systems. Most prominently, dynamic adaptations of workflows to new workflow specifications is regarded as an important functionality to enhance flexibility of workflow management systems.

The remainder of this thesis introduces the formal foundation, conceptual design, prototypical implementation, and the usage of dynamic adaptations of running workflow instances. It is not surprising that dynamic adaptations impose considerable requirements on the design of workflow languages as well as on the theoretical basis, conceptual design, and implementation of workflow management systems. By using correctness criteria based on formal specifications of workflows, dynamic adaptations can be controlled properly. Today, flexibility issues in general and dynamic adaptations of running work-flow instances in particular is an important research topic in workflow management, and the results presented in this thesis will contribute to this research area.

Integration, Re-Use, and Software Components

Organizations typically operate in complex technological and organizational environments. In the technical environment, for instance, organizations may rely on a variety 10 of different information systems to store, retrieve, manipulate, and manage application specific data. Some of these systems may use relational database technology, while others may use pre-relational systems. As a typical pattern, the technical and organizational environment of a company has gradually evolved since the company was founded. Often the mission-critical application systems of a company are the oldest and most reliable ones, because the company was founded to perform these operations. These application systems are known as legacy systems. Due to the high risk and considerable resources involved, re-implementing the software systems at the core of a business is a task often avoided. Not surprisingly, these applications have typically been developed without process orientation or workflow management in mind. In particular, they often lack the flexibility which is required to adapt applications quickly and with little effort to new requirements as they emerge. Workflow management systems are used to provide this functionality. Therefore, workflow management systems have to communicate with these applications to support mission-critical business processes.

The integration of legacy systems in workflow applications is one of the most important issues in workflow management from a practical perspective. Its importance comes from the mission-criticality of legacy applications. As a result, not only the legacy applications have to provide high performance and availability, but also their integration in the workflow application. Approaches to integrate legacy applications include wrapping techniques, in which the legacy system is wrapped with a software layer, called wrapper. The wrapper hides the particularities of the legacy system and presents a clean interface to the workflow management system. This interface is used to invoke legacy applications, to transfer input data to and to get output data from these systems.

Another requirement of business applications is the fast development of new services or the fast improvement of existing ones. The approach of creating new services by assembling processes or process fragments leads to faster design and implementation of these services. In this context, the presentation of services to customers using Internet technology is important, since new services can be made available instantaneously. The fast and convenient creation of new application processes and their implementation using workflow technology requires support on two areas: First, re-use of existing workflow specifications in the context of multiple applications. Second, the convenient integration of application programs in workflow applications. Issues related to these areas are discussed in turn.

In order to create new workflows from existing ones, workflow specifications have to be re-used in different workflow specifications. A formal specification of workflow schemas and how workflow schemas can be assembled is a prerequisite to support work-flow schema re-use. A workflow specification library can be build as a repository for application processes and their modeling and implementation as workflows. A work-flow specification library can also be used to create new workflows from fragments of existing ones, as discussed above. This approach is closely related to component-based software development: Workflow specifications are regarded as software components, and by assembling workflow components, new workflow specifications can be designed conveniently. The introduction of new services by creating new application processes cannot be achieved without adequate technological support.

In this context, application frameworks are a promising approach. These frameworks aim at efficient development of complex software system by providing basic functionality and defined ways to extend it. Instead of starting from scratch, the system development starts from a set of predefined classes, which are then extended to suit the needs of a particular application. Frameworks are typically based on object technology, since abstraction concepts provided by the object paradigm — for instance, inheritance, aggregation and polymorphism — prove useful in this context. This approach aims at developing objects and collections of objects which have a meaning in the application domain. These objects are known as business objects. In a Request For Proposals for a Business Object Facility [77] issued by the Object Management Group (OMG), business objects are defined as follows: A business object is defined as a representation of a thing active in the business domain, including at least its business name and definition, attributes, behavior, relationships, rules, policies and constraints.

A business object may represent, for example, a person, place, event, business process or concept. Typical examples of business objects are: employee, product, invoice and payment. The business object [. . .] is implemented by one or more objects in the information system.

Business objects and business-object frameworks are also useful with respect to the integration of domain-specific applications, using wrapper-techniques. In addition, software components in general and business objects in particular are useful for fast and efficient development of new application systems or improvement of existing ones [104]. Since in contemporary businesses, new services can often be

implemented by new applications or new business processes, component-based software development is a promising approach.

Business objects can be re-used as software components in a variety of applications. Hence, software development and maintenance can be simplified considerably using business objects. The advantages of software components can be exploited even more when they are integrated by a workflow management system, which adds process support to business objects. To go one step beyond, workflow functionality can also be designed and implemented by specific business objects. This approach integrates the application specific functionality and workflow functionality in a common framework. In this thesis, an object-oriented approach to the conceptual design and the implementation of workflow management systems is proposed, which simplifies the integration of workflow functionality, software components, and application frameworks. Concepts, architectures, and upcoming standards in the context of business objects and business-object frameworks are discussed in [112].

Availability, Scalability, and Distributed Workflows

High availability is an important aspect of workflow applications, often characterized by 24/7 availability, meaning that the applications have to be operational 24 hours a day, 7 days a week. This means that down times due to system administration activities are also not acceptable. In the workflow context, installing new or modified workflow specifications or changing the organizational and technical environment of the workflow application have to be made while the system is operational. Safe operation and fault tolerance are related features in this context. This means that in the event of a system failure, after the system restarts, it has to be recovered to a consistent state. In the database context, similar requirements have led to the transaction concept, involving synchronization of parallel transactions and automatic recovery after system failures [7, 37]. In particular, after a system failure, the database will be restored to a consistent state using restart procedures. In the workflow context, at each point in time the current status of the workflow management system has to be stored persistently. Persistently stored workflows are therefore a prerequisite to recover the current state of the workflow management system after a system failure has occurred.

Recently there is an increasing interest in application processes involving multiple organizations. In particular, workflow management is seen as an important enabling technology to facilitate workflows between multiple enterprises. This emerging field of work-flow management is characterized by the term inter organizational workflow. From a technological perspective it is not reasonable to control inter organizational workflows in a single dedicated machine. This would lead to a situation in which the internal workflows of one organization are controlled by a workflow server run by another organization. Most likely, acceptance issues would occur in this context which would make this approach not feasible. Distributed workflow executions can be motivated by inter organizational work-flows, in which all participating organizations have the opportunity to control the parts of the overall workflow, for which they are responsible. Due to the fine granularity of work-flow execution control, group-wide workflows can be controlled on a local machine, while department-wide workflows may be executed on a dedicated workflow server running under the control of that department. Finally, inter organizational workflows can be implemented conveniently by placing the relevant parts of the global workflow to dedicated machines of the organizations involved. Hence, local autonomy, which is an important issue from an organizational point of view, is supported by the software infrastructure.

Since workflow management aims at supporting mission-critical application processes, a large number of concurrently executed workflows can be expected. Hence, the workflow management system has

to cope with a heavy load, which is characterized by a large number of workflow activities and many persons involved, i.e., the workflow management system has to provide scalability, meaning that the system can perform work- flows faster or more workflows can be performed if additional resources are installed. As a consequence, workflow applications have to be scalable. Adding resources to a system which uses a fully distributed approach improves the efficiency and enhances the throughput, provided additional communication cost incurred by additional resources does not hamper these advantages.

Workflow Application Development Processes

Workflow applications are developed in workflow application development processes, as specific software development processes, in which typically numerous persons with different backgrounds and experiences are involved. Software development processes are studied in software engineering, and workflow development processes are identified in this thesis as software development processes with a number of specific properties. To cope with the properties of workflow development processes, these processes are studied, and a workflow development methodology is introduced, aiming at assisting work- flow project participants in planning and conducting workflow application development projects.

Research Issues not Addressed

While this thesis addresses a number of important research issues in workflow management, there are other important research issues which are not addressed; some important research directions in workflow management are sketched in the remainder of this section. We remark that this section does not discuss related work (an extensive analysis of related work will be given in Part III). In contrast, research topics are briefly sketched, and references to important contributions addressing the respective topics are provided. The reader may consult these references for more information.

One area of workflow management research deals with architectures of workflow management systems, based on different goals. One goal is the development of workflow servers according to particular aspects of workflows [56]. Other approaches in workflow management architectures consider partitionings of workflow servers in compliance with the organizational and technical structures of an organization [117, 116] as well as typical access patterns to reduce communication overhead, using role information [4, 85]. In [34], the use of queuing theory for the partitioning of workflow servers is discussed. The integration of workflow technology and computer supported cooperative work is an important research issue in the workflow context [59]. By integrating methods and tools for supporting collaborative work with methods and tools for modeling and controlling the execution of activities, advanced information systems can be developed. Taking advantage of a strong background in database systems, a number of research projects deal with transactional workflows and transactional processes [93, 99, 89]. Approaches in this context aim at defining execution guarantees for workflows, just like execution guarantees for application programs in the context of database transaction processing. Active databases are another background for research in workflow management. The application of methods and techniques from the area of active databases are twofold. First, workflows can be modeled using so called event-condition-action rules, a basic formalism in active databases. Second, active databases can be used as implementation platforms for work- flow management systems [33]. The combination of workflow technology and product data management is an interesting approach for supporting complex application processes in engineering with workflow functionality [70].

Only recently, another application area of workflow management came up: electronic payment. In this context, advanced security and authentication mechanisms are used to make sure payment is done in a reliable and safe manner; workflow management systems can be used for modeling and controlling the steps that occur during an electronic payment [2, 94]. Temporal aspects are another research issue in workflow management. The main idea of work in this area is modeling the duration of workflow activities explicitly; there are time intervals for each workflow activity, during which it has to be executed [25]. Finally, migrating workflows is mentioned as a research issue in workflow management. The basic idea in migrating workflows is that workflows can migrate to particular sites of a given organization or even to sites maintained by other organizations [90], with the aim to make use of local knowledge and expertise, and as a means to provide load balancing. Approaches in the context of migrating workflows are related to agent technology, in which autonomous agents with defined goals get in contact with other agents in order to negotiate and, finally, to reach their goals.

workflow 管理系统：

正式奠基，概念设计，实施问题

workflow 管理的目的在于建模和控制应用程序的执行动态和异构的组织和技术环境正如事实 [67, 32]。因此，对于 workflow 技术的动力来自的具体要求信息系统应用。要为 workflow 提供了一个简洁的管理动机换货和一般的主题本文特别，特殊性能解决信息系统的引入的应用程序。基于这些性质，重要的在 workflow 管理问题研究推导，和工作的正式基础流管理被引入，作为用于解决这些问题的基础。关键在计算机科学方面发展的理论成果，并显示其有效性在操作原型软件系统。基于相对于理论成果要解决的研究课题，本文探讨了 workflow 的概念设计管理系统。引入的概念和方法的原型进行验证 workflow 管理系统的实现。

背景，动机和研究问题

21 世纪带来了如何组织变化剧烈在商业，公共管理，教育，科学和工程工作，如何与其他组织和他们的客户进行互动。许多这些新的发展是由最近的技术进步推动，例如是否可用的计算和网络资源成本低，最突出的是，快速增长的互联网。其结果是，个人可以方便的现在与同事沟通朋友们在世界的其他部分，其特征在于地球村隐喻。在工商部门，技术进步创造了什么，现在被视为全球经济。

或许其中最明显的外观 developmentsis 在线的极大兴趣应用程序，即吸引互联网上的潜在客户，接受订单，交换产品规格或科学数据，并在一定程度上提供的在线交付货物，并通过互联网促进协商和支付。在这些情况下，产品和服务现在提供给客户世界各地瞬间之后他们被释放。在线应用此属性有强大的影响他们的设计实施，和维护信息系统支持这些应用程序。特别是，它激励用于快速创建新的服务和需要现有的灵活地适应。在技术层面上，系统故障不再 AC-接受的，因为它们立即暴露给客户，谁都有可能选择其他组织以提供所需的服务。如将在此剩余的动机章，这些要求强加在 workflow 管理有趣的研究问题，这可以通过使用各种计算机科学概念，方法和技来解决。在讨论论文和概念涉及的研究问题之前，方法和技术用于解决这些问题，在应用一些一般性意见 workflow 管理系统的背景是适当的。但从一个面向应用的角度，一个重要的方法来应对新要求的应用和 envi-不同方面的显式建模境中它们执行。在这方面最突出的个别方面是过程导向，这已成为被广泛接受和成功的方法描述工作是如何在组织中进行的。在这个过程中定位的方法，应用程序是关注的中心，而不是单独的函数在这些过程中形成的。为了激励过程导向，简单的介绍一下传统的方法模型工作的组织取。在传统的做法，工作被分解成细小的颗粒，称为功能。工作是围绕功能组织，和功能之间的关系，通常没有考虑到。尽管在事实功能由各个应用程序适当地进行，工作在这些设置往往是低效执行。对这一事实的原因包括多种媒体来存储应用程序特定数据，以及多余的工作和等待时间引入由各个功能之间的相互关系的管理不善。作为制造业实现了早期在那里生产线已推出几十年前，以提高制造业产品的生产，工作，通常是形成过程。这意味着，从个人活动之间的关系有要考虑到，和工作应围绕进程，而不是被组织周围的各个功能。这一观察也适用于该管理 IN-流程形成而不是实物商品;这些过程被称为信息处理。

信息流程中出现上面提到的域。在商业组织系统蒸发散，保险索赔管理和信用请求处理都是典型的例子应用程序进程。在公共管理部门的示例应用程序以流程 ING 的提议兴建私人住宅。在一所大学的学生入学涉及了一些步骤，它可以在教育被表征为一个应用进程

部门。实验室信息管理,实验管理,以及建筑复装置的灰是在科学中的应用过程的典型实施例和工程。进程的显式建模在其中首次出现的应用程序上下文商业机构;在这些设置应用程序进程被称为业务。正如事实,业务流程的概念是由锤与钱皮[42]为“定义的该接受输入的一种或多种和活动集创建的输出即是价值给客户。”因此,业务流程通常涉及关键的一个组织正如事实。任务紧迫性的一个很好的迹象是,该组织成立执行这些过程[68]。业务流程建模的活动被称为是业务流程建模。值得一提的是,业务流程建模并不限于在商业机构的应用进程。相反,它可以还可以应用到在其中应用进程需要进行建模,对其他领域的比如在公共管理,教育,科学和工程。抽象-从特定域ING,术语应用程序和业务流程可以被使用。业务流程模型的设计通常是开发按业务领域的专家,而不是由计算机科学家。问题相关业务过程建模因此将不本文予以解决。经过一个组织已经占据其业务流程,软件开发都面临着如何正确地支持这些进程的决定。在此刻工作流技术进入画面。如上所述,工作流管理旨在控制业务流程的自动化部件的执行。因此,它提供的形式化表达业务流程和的自动部件它们所执行的技术和组织环境。在一般情况下,这些形式主义称为工作流语言。使用工作流语言,工作流规范被创建,它被用作工作流管理系统的输入。因此,工作流语言必须正式,从而使工作流管理系统可以根据给定的工作流规范明确地控制工作流。基于信息的系统应用的特定要求,分析如在文献中报道的概念性工作,和现有的作品而言局限性流量管理系统,研究领域中的工作流管理和相应的工作流管理系统的功能进行了说明:

灵活性和动态改编:虽然灵活性一直是一个主要的工作流技术从一开始,只有有限的功能对于灵活的工作流管理由传统的工作流提供管理系统。今天,更多的灵活性需要被广泛接受在工作流的社区,由若干最近车间[101, 61 表示, **集成,再利用,以及软件组件:**重新使用的工作流程规范的系统蒸发散组装新的工作流程规范,从现有的一个重要在流程管理方面。工作流程规范的正式基础,在这方面的一个重要的先决条件。在一个更为技术的层面,融合外部应用程序和现有软件组件是一种很有前途的方法。**可扩展性,可用性和分布式工作流:**可扩展性和可用性具有重要的方面由信息系统来满足,指示通过安装额外的资源,以提高吞吐量和分别服务无故障。分布式工作流管理是一个有前途的方式,加强工作流程的可伸缩性和可用性管理彪系统,如将在下面详细讨论。在本节的剩余部分,这些研究领域的动机在一些细节上,研究发现存在问题的,适合的工作流程支持的概念和技术要求端口导出。

灵活性和动态改编

鉴于当今的组织在电子商务的动态结构,公共管理,教育,科学和工程,这是不可能的申请过程是一旦模型将没有任何变化重复执行。反之,流程可能随时间而逐渐改变,以反映在该方法的环境的变化-流程发展。当然,弹性不限于软件系统。根据到字典中,神器是灵活的,如果它可以“很容易被改变,以适应新的条件。”工作流管理的灵活性用于与此含义:工作流应用程序是柔性的,如果它可以很容易地改变,以适应新的条件下,通过在改变施加在应用过程中的环境。根据不同的应用领域下,有可能是快速发展的应用程序和有可能慢慢地不断发展的。在某些应用领域甚至有可能是哪些没有流程更改所有,例如当工艺结构是由法律法规规定的。预计不会长时间切换。对于这些应用程序进程传统工作流的支持就足够了,对于别人,也不会。如上所述,灵活性一直工作流管理的一个主要的动机从开始:通过显式建模过程和各种附加方面将在本文后面讨论,工作流程规范单个尺寸可如有变更,不改变其他方面。如将在下面所指出的,工作流的语言被设计成,允许执行修改某些模块化方式

方面,同时保留其他方面。在一般情况下,工作流管理系统提供执行修改的灵活性工作流规范,以及未来所有的工作流程将使用新的,改进的工作流规范。然而,工作流规范给定的工作流不能在传统的工作流管理系统改变。这意味着,一旦某作品流已经开始,没有变化可以应用于其结构。这是一个严重的限制工作流管理系统,工作流程,因为可以长期运行的活动,可能,或者甚至几个月或几年运行几天,几周。限制修改新工作流程将使工作流技术在这些设置几乎无法使用,因为长期

运行的工作流程将不得不继续以取代工作流程规范为长时间。这将是更足以使运行的工作流要适应新的和改进的工作流程的规范。因此，要求快速反应以在一个过程环境的变化不是由位的功能相匹配传统的工作流管理系统的性。最明显的是，动态调整工作流程，以新的工作流规范被认为是一个重要的功能提高工作流管理系统的灵活性。在本文的其余部分介绍了正式奠基，概念设计，实施和运行的工作流 IN-的动态修改的使用立场。它是不奇怪的动态适应施加相当大的要求工作流语言的设计以及理论基础上，签署和实施工作流管理系统。通过使用正确基于工作流程的正式规范 RIA，动态的调整可以控制正常。在运行的作品而言，一般和动态调整的今天，灵活性问题流程实例特别是在工作流管理的重要研究课题，并本文介绍的结果将有助于这一研究领域。

集成，再利用和软件组件

组织通常在复杂的技术和组织环境工作。在技术的环境中，例如，组织可能依靠各种不同的信息系统来存储，检索，操作和管理应用程序-具体数据。一些系统可能使用关系数据库技术，而另一些可以使用前关系型系统。作为一个典型的模式，技术和组织公司的逐渐演变自公司成立。经常公司的关键任务应用系统是最古老和最可靠的人，因为公司的创立是为了执行这些操作。这些应用系统被称为遗留系统。由于高风险和相当多的资源参与，重新实现软件系统在业务的核心是一个任务避免。不足为奇的是，这些应用程序通常被开发，而不记住流程导向或工作流管理。特别是，他们通常缺乏这是需要快速且毫不费力地以新的适应应用的灵活性要求他们出现。工作流管理系统来提供此功能。因此，工作流管理系统与这些通信应用程序以支持关键任务的业务流程。遗留系统的工作流应用程序的集成是一些最重要的一个在工作流管理问题，坦从实用的角度来看。它的重要性来从遗留应用程序的任务关键性。其结果，不仅是传统的并发症必须提供高性能和高可用性，而且他们在整合工作流应用程序。方法集成遗留应用包括包装技术，其中，遗留系统是包裹着一个软件层，称为包装。包装器隐藏了旧系统的特殊性，提出一个干净的界面工作流管理系统。此接口用于调用传统应用程序，输入数据传送到并获得来自这些系统的输出数据。业务应用的另一个要求是新业务的快速发展或者将现有的快速改善。通过 AS-创建新服务的方法工艺或流程片段导致更快的设计与实现这些服务。在这种情况下，服务于客户的呈现使用因特网技术很重要，因为可以使新的服务提供瞬间。该快速便捷的创建新的应用程序及其实施使用工作流技术，需要在两个方面的支持：一是重新使用现有的工作流程规格中的多个应用程序的上下文。其次，方便集成在工作流应用程序的应用程序化。与这些领域有关的问题是反过来讨论。为了从现有的创建新的工作流程，工作流程规格有在不同的工作流规范被重新使用。工作流程的正式规范模式和工作流程如何架构可组装是支持工作流的先决条件流架构重新使用。工作流程规范库，可以建设成为一个资源库。申请流程及其建模和执行情况的工作流程。一个工作-流规范库也可用于创建从片段新的工作流程现有的，如以上所讨论。这种方法是密切相关的基于组件的软件开发：工作流程规范被视为软件组件，并通过组装工作流程组件，新的工作流规范可设计方便。通过创建新的应用程序推出新服务没有足够的技术支持是无法实现的。在这种情况下，应用程序框架是一个有希望的方法。这些框架通过提供基本功能着眼于复杂的软件系统的高效发展和定义的方式来扩展它。相反从无到有，系统开发的开始从一组预定义类，这是开始然后扩展到适应杆的需要满足特殊的应用程序。框架通常基于对象的技术中，由于抽象由对象提供的范式化的概念-例如，继承，聚集和多态性-证明在这方面是有用的。这种做法的目的是发展转播项目与具有在应用领域意义对象的集合。这些对象称为业务对象。在招标信息为业务事先获拍摄对象设备[77]由对象管理组织（OMG），业务对象发行定义如下：业务对象定义为活跃在事情表示内斯域，至少包括其企业名称和定义，属性，行为，关系，规则，策略和约束。业务对象可表示，例如，一个人，地点，事件，业务流程或概念。业务对象的典型实例是：雇员，产品，发票和付款。业务对象[...]是由一个或多个对象执行在信息系统中。业务对象和业务对象框架也是有用的相对于所述英特特定领域的应用，利用包装的技术格雷申。此外，软件特别是在一般和业务对象的组件是快速，高效实用新的应用系统或现有的[104]的改善发展。以来在当

代的企业，新的服务通常可以通过新的应用实施或新的业务流程，基于构件的软件开发是一种很有前途在各种应用中的业务对象可以被重新用作软件组件。因此，软件的开发和维护能够大幅度使用内斯对象。软件组件的优点是可以更时被利用它们由 workflow 管理系统，这在增加了工艺集成支持业务对象。超越一步，workflow 功能还可以设计并通过具体的业务对象实现。这种方式整合了应用程序 - 具体的功能和 workflow 功能，在一个共同的框架。在本文中，面向对象的方法，概念设计和 workflow 的实施管理系统，提出了一种简化了 workflow 功能 - 的积分先进而精湛，软件组件和应用框架。概念，结构，和在业务对象和业务对象的框架下即将推出的标准在 [112] 进行了讨论。

可用性，可伸缩性和分布式 workflow

高可用性的 workflow 应用的一个重要方面，特点往往是 24/7 可用性，这意味着应用程序必须是操作一天，7×24 小时一周天。这意味着，下降时间由于系统管理活动也不能接受的。在 workflow 的背景下，在安装新的或修改的 workflow 规范工作，或改变 workflow 的组织和技术环境的应用化必须做在系统正常运行。安全操作和容错在这方面相关的功能。这意味着，在系统发生故障的情况下，后系统重新启动，它必须恢复到一致状态。在数据库方面，类似的要求已经导致了事务的概念，涉及的同步并行事务后系统故障 [7, 37] 自动恢复。尤其是，系统出现故障后，该数据库将利用亲重启恢复到一致状态。在 workflow 的背景下，在每个时间点的工作流的当前状态管理系统必须持久地存储。持久存储 workflow 是脱颖而出的先决条件后恢复 workflow 管理系统的当前状态发生了系统故障。最近有在应用进程的兴趣日益增加涉及多个组织。特别是，workflow 管理被看作是一个重要的有利技术，以方便多个企业之间的工作流程。这一新兴的作品而言领域流管理的特征在于由术语组织间的工作流程。从技术角度来看它是不是合理控制组织间的工作流程一个专用的机器。这将导致一种情况，其中，内部的工作流一个组织通过由另一组织运行 workflow 服务器控制。最可能的是，接受的问题可能发生在这种情况下这将使得该方法不可行。分布式 workflow 执行可以通过组织间 work-动机流动，其中所有参与组织有机会控制的部分整体工作流程，为他们负责。由于作品而言，细粒度流程执行控制，整个集团的工作流程可以在本地计算机上进行控制，全系 workflow 可能在一个专门的工作流服务器的运行被执行该部门的控制之下。最后，组织间的工作流程可以被立即通过将全局 workflow 的相关部分方便地执行完成到专用有关组织的机器。因此，地方自治，这是一个重要从组织的角度来看问题由软件基础设施的支持。由于 workflow 管理旨在支持关键任务应用程序，可以预见大量的并发执行的工作流程。因此，workflow 管理系统具有应付重负载，其特征在一个大量的 workflow 活动，并牵涉到很多的人，即 workflow 管理系统必须提供的可扩展性，这意味着该系统可以执行 workflow 如果安装了额外的资源可以进行更多的工作流程更快或。因此，workflow 应用程序必须是可扩展性。将资源添加到系统它使用一个完全分布式的方法提高了效率，并提高了吞吐量由额外资源投入产生的，提供了额外的通信费用不妨碍这些优点。

workflow 应用程序开发流程

workflow 应用程序的 workflow 应用程序开发流程开发，如具体的软件开发流程，通常很多人与昼夜温差同的背景和经历都参与。软件开发过程是研究了软件工程和 workflow 开发过程在确定本文为软件开发与一些特定属性的处理。至应付 workflow 开发过程的属性，这些过程都和工作流开发方法的介绍，旨在协助 workflow 工程参与策划和实施 workflow 应用程序开发项目。

研究问题未解决

而本文中的 workflow 管理解决了许多重要的研究课题包换，还有哪些问题不解决，其他重要的研究问题；一些重要在 workflow 管理研究方向勾勒在本章节的其余部分化。我们注意到本节不讨论有关工作（的广泛分析相关工作将在第三部分给出）。相反，研究课题进行了简要勾勒，并提供解决各自的主题作出了重要贡献引用。读者可以了解更多信息请咨询这些引用。

workflow management research的一个领域与工作流程的架构涉及管理系统，是根据不同的目标。一个目标是 workflow的发展根据 workflow[56]的特定方面的服务器。在工作流程的其它方法管理架构考虑 workflow服务器的多种分割符合一个组织[117, 116]的组织和技术结构以及典型的访问模式，以降低通信开销，使用角色信息[4, 85]。在[34]，使用 workflow服务器的分区排队理论的讨论。 workflow技术和计算机支持的协同工作的集成是一个重要的研究课题在工作流程方面[59]。通过整合的方法和工具用于支持与方法和工具协同工作进行建模和控制活动的执行，先进的信息系统可以开发。AD-服用在数据库系统强大的后台华帝，一些研究项目涉及事务性的工作流程和交易流程[93, 99, 89]。接近这背景瞄准界定执行担保工作流程，就像执行瓜拉尼开杆在数据库的事务处理的上下文中的应用程序。 活性数据库是在 workflow管理研究的另一背景。应用程序从活动数据库的区域的方法和技术是双重的。首先，工作流程可以使用所谓的事件 - 条件 - 动作的规则，在活性碱性形式主义被建模数据库。其次，活性的数据库可以用作：为工作实施平台流量管理系统[33]。 workflow的技术和产品的组合数据管理是支持复杂应用处理一个有趣的方法在工作流功能[70]工程。只是到了最近， workflow管理的另一个应用领域上来：电子付款。 在这方面，高级的安全和认证机制用于确保付款以可靠和安全的方式进行； workflow管理系统可以电子缴付一期间用于建模和控制所发生的步骤[2, 94]。时间方面是 workflow管理的另一个研究课题。该在这方面的主要思想是明确的建模 workflow活动的持续时间；有时间间隔为每个 workflow活动，在此期间，它必须被执行[25]。最后，迁移 workflow被提及作为 workflow管理的一个研究课题。

在迁移 workflow的基本思路是， workflow可以迁移到特定的网站一个给定的组织或甚至其他组织[90]保持位点，其目的充分利用当地的专业技能，并提供负载手段。在迁移 workflow的情况下方法都涉及到代理技术，其中明确目标自治代理获得与其他代理商联系，并最终的目标。