

**本科毕业论文(设计)**

**题目：**实时气象大数据分析与预测

**学 院** 管理科学与工程学院

**专 业** 信息管理与信息系统

**班 级** 2014级2班

**学 号** 20140614107

20140614234

20140614204

**姓 名** 付腾达 王丽 程前

**指导教师** 张贵杰

山东财经大学教务处制

二Ｏ 一八 年 四 月

**山东财经大学学士学位论文原创性声明**

本人郑重声明：所呈交的学位论文，是本人在导师的指导下进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在论文中作了明确的说明并表示了谢意。本声明的法律结果由本人承担。

学位论文作者签名：

年 月 日

**山东财经大学关于论文使用授权的说明**

本人完全了解山东财经大学有关保留、使用学士学位论文的规定，即：学校有权保留、送交论文的复印件，允许论文被查阅，学校可以公布论文的全部或部分内容，可以采用影印或其他复制手段保存论文。

指导教师签名： 　 论文作者签名：

　　 年 月 日　　　 　　 年 月 日

实时气象大数据分析与预测

摘 要

随着信息化时代的进步，气象分析与预测越来越与人们息息相关。从上世纪二、三十年代开始，传统的无线电技术兴起，可以将高空中观测到的气象数据传递给地面，由此人类可以获取到很多观测资料，通过对这些数据的简单处理，就可以进行气象预测。对这些数据进一步统计处理分析，并根据流体学、动力学等物理学的原理，从而构建起一套基本的气象学科框架，并随着时代的进步，逐步对其进行完善。因此，气象学科是建立在观测资料的基础上，并由气象学家通过各种先进的统计方法进行深入的科学分析研究，发现其中的内在机理，进而形成最终的一套完善的气象学科理论体系。随着资料的不断丰富完善，数据越来越大，人工智能的兴起，单台计算机已经不足以处理如此众多的气象数据，分布式存储、分布式计算逐步的发展，逐步也引入到了气象学科。

关键词：实时、气象、大数据、分析、预测。

Financial Wallet Platform

ABSTRACT

With the progress of the information age, meteorological analysis and prediction are more and more closely related to people.From 20 to 30 s of the last century, the traditional radio technology, the observed meteorological data can be high up in the air is passed to the ground, the man can get a lot of observation data, by means of the simple processing of the data, you can proceed to the weather forecast.Further statistical processing of the data analysis, and according to the fluid study, dynamics, such as the principle of physics, so as to build up a basic framework of meteorological science, and make progress with The Times, to improve step by step.Meteorological science, therefore, is established on the basis of observation data, and by meteorologists through a variety of advanced statistical methods in-depth scientific analysis and research, found that the internal mechanism, and finally

a set of perfect weather discipline theory system.Constantly enrich and perfect as the data is more and more big, the rise of artificial intelligence, a single computer has not enough to deal with so many meteorological data, distributed storage, the development of distributed computing, gradually, gradually also introduced to the weather.

**Keywords**：runtime、weather、bigdata、analyse

**目 录**

1. 引言 i
2. 市场分析 i
3. 设计目的 i
4. 设计目标 i
5. 项目技术应用 i
6. 系统分析 ii
7. 系统工程时间预算 iii
8. 需求分析 iii
9. 系统生命周期图 iii
10. 系统用例 iv
11. 系统设计 v
12. 系统功能详细分析 v
13. 系统组件图 v
14. 系统流程图 vi

整体： vi

数据端： vi

1. 系统实施 vii
2. 系统测试 vii
3. 系统效果图 viii
4. 系统工程决算 viii
5. 结束语 viii
6. 致谢 ix

参考文献 x

山东财经大学本科毕业论文（设计）中期检查表 xi

山东财经大学本科毕业论文（设计）开题报告 xii

山东财经大学本科毕业论文（设计）评阅审核表 xiv

1. 引言
2. 市场分析

目前我国互联网平台处于高速发展期，互联网+、大数据这些关键字称谓互联网界的热点，

1. 设计目的

本系统

1. 设计目标

实现一个基于大数据的气象大数据平台，提供实时气象状况查询、历史气象查询、气象分析以及气象预测等功能。本项目将分四部分组成——数据端、后端、前端、测试。

1. 项目技术应用

数据架构：hadoop+spark

数据组件：hdfs+spark+spark-streaming+kafka

数据技术栈：

1：离线大数据储存——Hadoop

Hadoop是一个可扩展的、可靠的分布式的大数据储存与计算框架，这使得它成为当下云平台架构下的首选。Hadoop主要由两部分组成——HDFS和MapReduce，在本篇文章中，只用到了HDFS。HDFS的高容错、高可靠性允许用户将Hadoop部署在低廉的硬件上，从而搭建起分布式系统，轻松利用集群的性能完成海量数据的处理。

在本项目中，Hadoop的作用主要体现在分布式储存上，即储存我们的原始气象数据集。

2：快速大数据分析——Spark

Hadoop和Sparkl联合构成了当今的大数据世界，Hadoop主要负责数据储存和资源管理，Spark负责一体化、多元化的不同规模的数据计算。Spark是一个通用的大规模数据快速处理引擎，相比于传统的以Hadoop为基石的第一代大数据技术生态而言，Spark无论是性能还是方案的统一性都具有极为显著的优势。

本项目中，Spark的主要作用体现在各种气象指标的分析。

3：高性能消息中间件——Kafka

Kafka是一个分布式的发布-订阅消息系统，具有高吞吐、低延迟、可扩展性、持久性、可靠性、容错性、高并发性等特点，主要由话题Topic、生产者Producer、消费者Consumer、服务代理Broker组成。生产者将数据发送到Broker代理，Broker代理有多个话题topic，消费者从Broker获取数据。

本项目中，Kafka的作用主要是模拟真实情况下，实时数据的生产与消费。

4：实时数据分析——Spark-Streaming

Spark-Streaming是Spark系统中用于处理流数据的分布式流式处理框架，能够以最低500ms的时间间隔对流数据进行处理，处理的延迟大概在1s，是一个准实时处理系统。并且，Spark-Streaming是一个非常出色的容错系统，具有极为高效的错误处理和恢复能力。

本项目中，Spark-Streaming的作用主要是实时消费Kafka中的气象数据。

后端架构：nodejs+redis

后端组件：express

后端技术栈:

1：高性能内存数据库——Redis

Redis是一个Key-Value内存数据库，支持string、list、set、zset等多种value类型，这些数据类型都支持push、pop、add、remove等操作，且这些操作都是原子性的。Redis还支持从简单到设置的主从异步复制，具有非常快的非阻塞的第一次同步，在网络分割上自动重新连接和部分重新同步。

本项目中，Redis的作用主要体现在储存实时产生的气象数据

2：事件驱动非阻塞服务器——Node.js

Nodejs是一个基于Chrome V8 JavaScript引擎的JavaScript运行时。nodejs使用一种事件 驱动的、非阻塞的输入/输出模型，使其轻量级和高效。nodejs的包管理npm，是世界 上最大的开源库生态系统。

本项目中，Node.js的作用主要体现在书写后端代码

3：web 开发框架——Express

Express是一个基于Node.js的web开发框架，它灵活、极简，可以帮助我们快速创建各

种APP和移动设备应用。同时它具有丰富的HTTP快捷方法和任意排列组合的Connect

中间件，让我们可以快速创建健壮、友好的API服务。 Express不对Node.js已有的特性

进行二次抽象，我们只是在它上面扩展了web应用所需要的基本功能

本项目中，Express的作用主要体现在提供web api接口

前端架构：

前端组件：

前端技术栈：

1. 系统分析

系统设计是对整个系统的结构的规划以及宏观控制的关键环节，主要是根据系统需求，系统的功能，用例等，分析应该设计什么模块，大体规划使用什么框架去实现，是进行系统评估的重要环节。前期的分析和设计非常重要，性能和优雅是个需要权衡的东西，前期的分析与设计就要设计好，优雅的设计可以减少后期很多工作，使得系统更容易实现低耦合高内聚的设计目标，同时要确保了代码对内部修改的关闭，和对扩展的开放，符合OO的设计思想，模块化管理。

1. 系统工程时间预算

系统的开发工作总体分做四个阶段：设计阶段，开发阶段，集成阶段和测试阶段。这四个阶段也可能交叉进行，比如开发阶段也会进行简单的单元测试，一个模块开发好了也会进行集成测试，中间开发也可能因为前期设计的不合理，而要进行重新设计。

在完成系统层级的开发后，鉴于测试工作量庞大而难以开展，因此测试方法采用小样本白盒测试和大样本黑盒测试的方式，并对测试结果进行相应的记录。

整体工程进度评估如下：设计阶段10日，开发阶段7日，集成阶段2日，测试阶段7日，工程评估26日。

1. 需求分析

本项目由四部分组成：数据端、后端、前端、测试，其中各自的需求分析如下：

数据端：

1. 获取原始气象数据，导入Kafka消息队列，准备消费

2. 使用Spark消费Kafka数据, 进行ETL转化, 转化成一份符合标准规范的日志, 并持久化

进Redis数据库

3. 使用Spark分析日志，计算各种指标

后端：

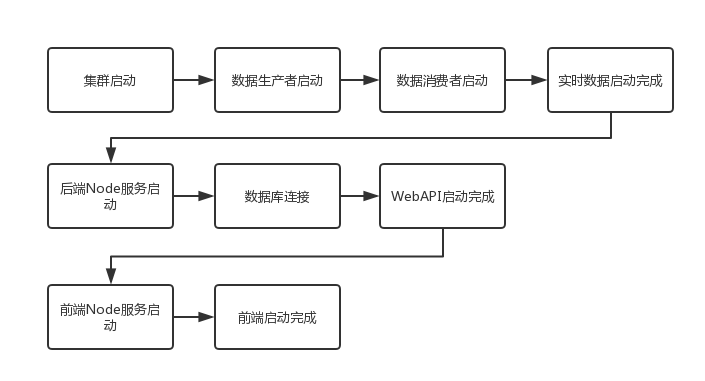
前端：

测试：

|  |  |  |
| --- | --- | --- |
| 部分 | 需求名称 | 进度 |
| 数据端 | 获取原始数据集 |  |
| 搭建大数据分析服务器 |  |
| 原始数据使用Shell脚本导入HDFS |  |
| 使用SparkStreaming模拟实时数据的产生并导入Kafka |  |
| 使用Spark消费Kafka数据，进行ETL转化，并导入Redis |  |

表2-1 系统需求表

1. 系统生命周期图

图2-1 系统生命周期图

1. 系统用例

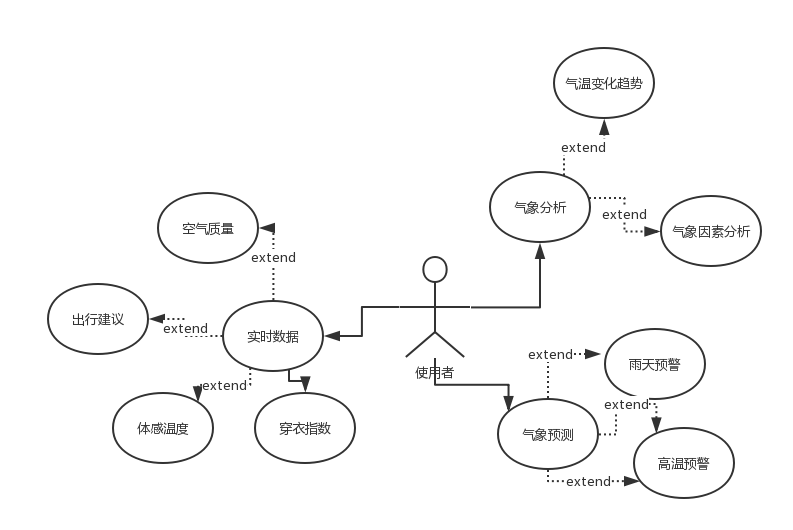


图2-2 系统用例图

1. 系统设计

功能流程是程序设计的重要工作环节，对于业务流程详细，合理的设计是确保系统稳定性，健壮性和友好性的重要保证。功能流程以用例规约表和活动图为主要设计手段，必要时以流程图为辅助。

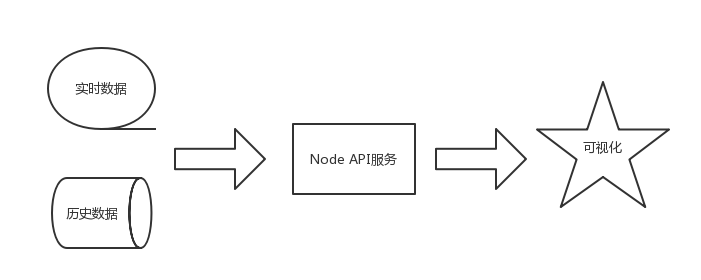
流程设计的设计应力求简明扼要，清晰准确的描述系统功能的流程结构。以下为本系统的功能流程设计图表。

1. 系统功能详细分析
2. 系统组件图

图3-1 系统组件图

1. 系统流程图

整体：



数据端：

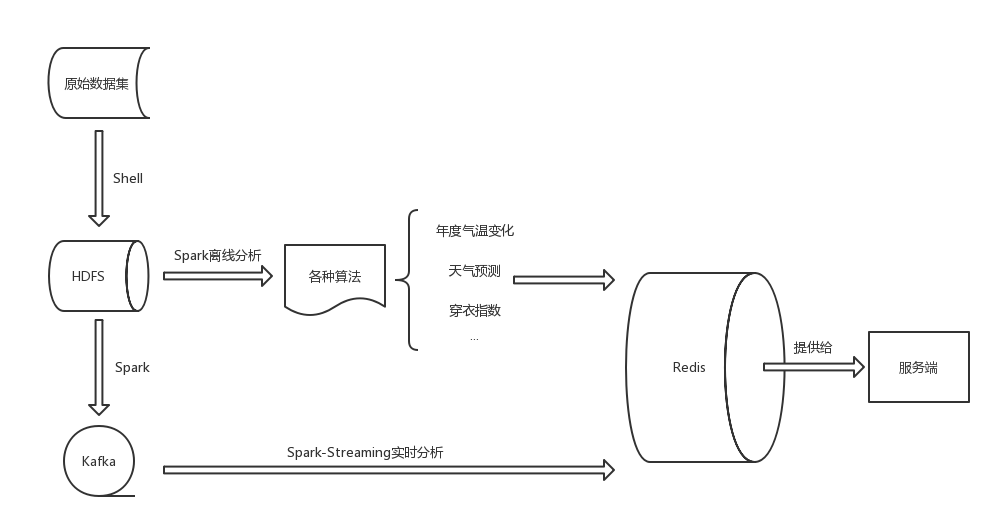


图3-3 系统数据端流程图

后端：

图3-4 系统服务端流程图

前端：

图3-5 系统前端流程图

1. 系统实施
2. 系统测试

系统测试是保证系统业务流程符合预期，系统健壮性符合设计要求的工作环节，也是系统开发流程中必须做的工作。本系统的系统测试工作分为黑盒测试和白盒测试两阶段，测试工作以黑盒测试为主，白盒测试为辅，主要通过RESTful风格的单元测试进行测试。因篇幅原因，此处不再对测试用例和测试案例进行展示，仅提供单元测试结果表和相应的单元测试结果数据。

测试工作预算：测试工作时间预算为测试用例评估1日，单元测试4日，黑盒测试3日，出具测试报告1日。

单元测试数据：系统单元测试共测试40个功能用例，编写测试用例119个，一次通过率87%，修复通过率11%，用例废弃率2%。

黑盒测试数据：黑盒测试共测试40个功能用例，一次通过率70%，修复通过率28%，用例废弃率2%。

白盒测试数据：白盒测试共测试40个功能用例，一次通过率90%，修复通过率8%，用例废弃率2%。

测试工作决算：测试工作时间决算为测试用例评估3日，单元测试6日，黑盒测试2日，出具测试报告1日。

本系统测试工作以确保系统可使用，保证系统健壮性和稳定性为目标。测试工作以单元测试为主要途径，黑盒测试和白盒测试相结合的方式进行测试。

整体代码质量参照以下代码质量缺陷分析报告和代码标准分析质量报告：

图4-1 代码质量缺陷分析报告

图4-2 代码标准分析报告

1. 系统效果图
2. 系统工程决算

工程预算和决算时间表如下：

设计阶段用例设计2日，流程设计5日，系统组件构架1日，需求复核评审1日，预算9日。实际用时用例设计5日，流程设计5日，系统组件架构2日，需求符合评审1日，实际12日，超标33%。

开发阶段系统架构搭建2日，技术评审1日，代码开发工作17日，预算20日。系统架构搭建2日，技术评审2日，代码开发工作26日，超标33%。

集成阶段前后端对接4日，系统覆盖率测试1日，出具集成报告1日，预算6日。实际用时前后端对接8日，系统覆盖率测试2日，出具集成报告1日，超标98%。

测试阶段测试案例设计1日，单元测试与白盒测试8日，出具测试报告1日，预算10日。实际用时案例设计3日，单元测试与白盒测试8日，出具测试报告1日，超标12日，超标20%。

总开发预算工时45日，实际66日，超标46%。

表4-29 系统工程决算表

|  |  |  |  |
| --- | --- | --- | --- |
| **系统工程预决算表** | | | |
| 工程阶段 | 预算数 | 决算数 | 超标值 |
| 用例设计 | 2 | 5 | 150% |
| 系统组件架构 | 1 | 2 | 100% |
| 需求符合评审 | 1 | 1 | — |
| 系统架构搭建 | 2 | 2 | — |
| 技术评审 | 1 | 2 | 100% |
| 代码开发工作 | 17 | 26 | 53% |
| 前后端对接 | 4 | 8 | 100% |
| 系统覆盖率测试 | 1 | 2 | 100% |
| 出具集成报告 | 1 | 1 | — |
| 测试案例设计 | 1 | 3 | 200% |
| 单元测试与白盒测试 | 8 | 8 | — |
| 出具测试报告 | 1 | 1 | — |

（单位：日）

1. 结束语

本系统以气象大数据平台为设计目标，截止至2018年4月25日工程结束本系统已经初步具备数据分析系统平台架构以及包含基本模块

通过张贵杰老师的指导以及吸纳学院各位老师中肯的意见和建议，通过本次毕业设计，我们不仅在需求分析，系统设计，系统编码以及系统测试等一系列系统开发的全生命周期流程都有更深刻的学习和认识，并且实践工作能力也得到了提升，对于书面知识有着更深刻的理解。同时，我们也对数据分析的需求和业务流程有一定的了解和认识。

1. 致谢

本次毕业设计的成功离不开学院以及指导老师的教诲与帮助，尤其张贵杰老师对我们的工作进行了深入而有见地的指导。同时本次毕业设计期间，我们也得到了学院许多其他老师的指导和帮助。本次毕业设计之所以取得我们预想的成果，与学院的各位老师的大力帮助密不可分，在此首先向给予本次毕业设计指导帮助的各位老师，特别是我们的指导老师——张贵杰老师表达诚挚的感谢！

四年来，在校园中我们感受到了学院领导以及各位导师的衷心投入与悉心栽培，四年的大学学习即将落幕，曾经遇到的困难和受到的指导鼓励仍然历历在目，因此我们大学四年得以全面成长，在理论、实践以及社会见闻上等方面获得全面的发展。相信我们走上工作岗位以后，我们也将继续承载着这份栽培与教诲，带着曾经在校园中收获的知识与感悟不忘初心，继续前行。

最后，再次向张贵杰老师以及为此次毕业设计提供各种支持帮助的学院老师们致以诚挚的感谢！

感谢各位老师的指导！

参考文献

[1] 朴灵.深入浅出Node.js [M]. 人民邮电出版社，2013（1）：10-188.

## 山东财经大学本科毕业论文（设计）中期检查表

学院：管理科学与工程学院 填表时间： 年 月 日

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 姓 名 | 付腾达  王丽  程前 | 学 号 | 20140614107  20140614234  20140614204 | 专 业 | 信息管理与信息系统 |
| 班 级 | 1402班 | 联系方式 |  | E-mail |  |
| 论文（设计）题目 | 气象大数据分析与预测平台设计与实践 | | | | |
| 已完成的工作： | | | | | |
| 未完成的工作： | | | | | |
| 计划完成时间和拟采取措施： | | | | | |
| 指导老师意见：  签名：  年 月 日 | | | | | |

## 山东财经大学本科毕业论文（设计）开题报告

学院：管理科学与工程学院 填表时间： 年 月 日

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 姓 名 | 付腾达  王丽  程前 | 专 业 | 信息管理与信息系统 | | 学 号 | | 20140614107  20140614234  20140614204 |
| 班 级 | 1402班 | 联系方式 |  | | E-mail | |  |
| 指导教师姓名 | 宋健 | 联系方式 |  | | E-mail | |  |
| 论文(设计)题目 | 气象大数据分析与预测平台设计与实践 | | | | | | |
| 计划完成时间 | 2018-04-27 | | | 计划撰写字数 | | 10k字 | |
| 本选题研究的目的、意义和实用价值： | | | | | | | |
| 本选题国内外研究现状： | | | | | | | |
|  | | | | | | | |

|  |
| --- |
| 围绕本选题已做哪些准备工作，计划再做的工作： |
| 拟解决的关键问题： |
| 主要参考文献： |
| 指导教师意见：  指导教师签名：  年 月 日 |

## 山东财经大学本科毕业论文（设计）评阅审核表

学院： 管理科学与工程学院 填表时间： 年 月 日

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 姓 名 | | 付腾达  王丽  程前 | 学 号 | 20140614107  20140614234  20140614204 | 专 业 | 信息管理与信息系统 |
| 班 级 | | 1402班 | 联系方式 |  | E-mail |  |
| 指导教师姓名 | | 张贵杰 | 联系方式 |  | E-mail |  |
| 论文（设计）  题目 | | 气象大数据分析与预测平台设计与实践 | | | 字数 |  |
| 指  导  老  师  意  见 | 评阅要点(主要包括：1.选题的理论和现实意义；2.作者运用所学知识分析问题和解决问题的能力；3.论文写作的规范性及文字表达能力；4.存在的问题；5.建议成绩)  指导教师： （签名）  年 月 日 | | | | | |

|  |  |
| --- | --- |
| 评 阅 老 师 意 见 | 评阅要点(主要包括：1.论文选题的实际价值与理论意义；2.论文的主要工作及成果；3.论文的文字表述、写作规范情况；4.论文的主要不足及修改意见；5.建议成绩)  评阅人： （签名）  年 月 日 |
| 答辩委员会（小组）意 见 | **该论文成绩综合评定为：**  答辩委员会（小组）负责人： （签名）  年 月 日 |
| 学 院 审 核 意 见 | 学院负责人： （签名）  （公章）  年 月 日 |