

注：教师姓名后留有一个空格，后面填写教师职称。下面加下划线。

阅后删除此文本框。

**本科毕业设计**

**房源数据分析与推荐系统设计与实现**

|  |  |  |
| --- | --- | --- |
| **学生姓名** | **：** | 黄浩星 |
| **学号** | **：** | 141544225 |
| **学院** | **：** | 互联网金融与信息工程程 |
| **专业** | **：** | 计算机科学与技术 |
| **指导教师** | **：** | 郭艺辉 **职称：** 讲师 |
| **提交日期** | **：** | 2018 年 2 月11 日 |

**本科毕业论文（设计）诚信声明**

本人郑重声明：所呈交的本科毕业论文（设计），是本人在指导老师的指导下，独立进行研究工作所取得的成果，成果不存在知识产权争议，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作品成果。对本文的研究作出重要贡献的个人和集体均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

 学生签名：

时间： 年 月 日

**关于论文（设计）使用授权的说明**

本人完全了解广东金融学院关于收集、保存、使用学位论文的规定，即：

1.按照学校要求提交学位论文的印刷本和电子版本；

2.学校有权保存学位论文的印刷本和电子版本，并提供目录检索与阅览服务，在校园网上提供服务；

3.学校可以采用影印、缩印、数字化或其它复制手段保存论文。

本人同意上述规定。

学生签名：

时间： 年 月 日

摘 要

最近几年来，大数据一词在人们的视线范围内出现得越来越频繁，大数据已经逐渐深入各个行业和业务领域之中，从信息爆炸时代产生的海量数据中分析，提取有价值的信息已经成为重要的生产要素。

房源数据在市场的交易中日益积累了海量数据，房地产市场交易的对象一般以房源作为基本单元，同时房源也是房地产的基本物理构成单元，房源数据主要包括：面积信息，周边环境，房价信息，地理位置，配套设施等。

如今人们越来越依赖于互联网信息，所以购房者在互联网的帮助下得到良好的服务质量的同时一定也留下了重要的数据痕迹，例如，兴趣取向，需求选择，环境考量等用户行为信息，当然，购买者同时也关心房源的历史发展信息，历史房价波动，房价升值等信息。这些动态的用户行为数据蕴含着巨大的商业价值，而且对房地产市场起着导向与反馈的作用，开发者可以从中挖掘丰富的数据信息从而拓展更大的市场信息。

大数据分析是用户交互，用户共识，信息交互之间的重要桥梁，大数据分析典型的应用就是通过海量数据检索，统计，比较，发现潜在的关联，然后获取关联的信息进行模式的预测和匹配分析。

**[关键词]**：大数据分析；用户行为分析；房源信息；用户推荐

Abstract

In recent years, the word BigData appears frequently in people's eyes. BigData has gradually penetrated into various industries and business fields. Analyzing and extracting valuable information from the massive data become an important factor of production.

In the market Housing transactions have accumulated huge amounts of data. Generally housing as the basic unit in the market transactions, while housing is also the basic physical units. Housing data include area information, the surrounding environment, price information, location information, facilities and so on.

Nowadays, people always depends on Internet information, therefore, home buyers, with the help of the Internet, get good service quality and leave important data traces while they buy housing in the Internet. User behavior information such as interest orientation, demand selection and environmental considerations are very important. Buyers are also concerned about the historical development information, fluctuations in historical prices, prices appreciation and other information. These dynamic user behavior data contain huge commercial value witch plays a guiding and feedback role in the market. Developers can explore huge data information to expand the market information.

BigData analysis is an important bridge between user interaction, user consensus and information exchange. The typical application of big data analysis is to find out the potential correlation through massive data retrieval, statistics, comparison, and then obtain the associated information to predict and match the patterns analysis.

**[Key Words]:** BigData analyzing; user behavior analyzing; housing data; user recommended

目 录

注意：（1）目录必须通过引用自动生成，不能手工生成；（2）论文修改后，重新生成目录，更新目录的标题及页码

样式：一级标题

字体：黑体，居中，加粗

字号：小三

段落：段前0行

段后0.5行

阅后删除此文本框

一级目录

字体：黑体、加粗

字号：小三

段落：1.5倍行距

阅后删除此文本框

[摘 要 I](#_Toc495071423)

[Abstract II](#_Toc495071424)

[目 录 III](#_Toc495071425)

[1 绪论 1](#_Toc495071426)

二级目录

字体：宋体

字号：四号

段落：1.5倍行距

阅后删除此文本框

[1.1 研究背景 1](#_Toc495071427)

[1.2主要研究工作 2](#_Toc495071428)

[2 相关技术 2](#_Toc495071429)

[2.1分布式系统 2](#_Toc495071430)

三级目录

注意：目录最多为三级

字体：宋体

字号：小四

段落：1.5倍行距

首行缩进2个字符

阅后删除此文本框。

[2.1.1 概述 2](#_Toc495071431)

[3 分布式移动 3](#_Toc495071432)

[4 结论 4](#_Toc495071433)

[参考文献 5](#_Toc495071434)

[致 谢 6](#_Toc495071435)

[附录 内容名称 7](#_Toc495071436)

注意：附录根据实际来确定是否需要

阅后删除此文本框

**分布式移动\*\*\*\*\*\*\*系统可靠性研究与设计**

正文

字体：中文为宋体：西文和数字等为Times New Roman

字号：小四

段落：两端对齐，首行缩进2个字符，1.5倍行距。

阅后删除此文本框

注意：封面题目，页眉题目，正文首页题目三者必须完全一致

字体：中文为黑体：西文和数字等为Times New Roman

字号：小二，加粗，居中

段落：段前为1行，段后1行

阅后删除此文本框

在20世纪50年代末到60年代初，计算机硬件从晶体管到集成电路，得到了飞速的发展，并朝着超大规模集成电路(VLSI )方向发展。

一级标题

基准样式：标题1

字体：中文为黑体：西文和数字等为Times New Roman

字号：小三

段落：左对齐

字号：小三，加粗

阅后删除此文本框

……

1 绪论

1.1 研究背景

二级标题

基准样式：标题2

字体：中文为宋体：西文和数字等为Times New Roman

字号：四号

段落：左对齐

阅后删除此文本框

在20世纪50年代末到60年代初，计算机硬件从晶体管到集成电路，得到了飞速的发展，并朝着超大规模集成电路(VLSI )方向发展。当时软件的重要性还不显著，尽管软件开发环境得以改善，如高级语言的产生，软件开发仍处于很低级的阶段，极大地依赖于开发人员的编程技巧，且主要关注的是软件的功能。但随着计算机硬件的飞速发展，软件需求量剧增，软件开发技术成为妨碍计算机进步的瓶颈。1968年在西德召开的国际软件工程会议上提出的“软件危机”的可靠性的重要性。据统计，计算机系统中，由于软件错误引起的故障占所有故障的65%。如表1.1所示。

注意：每个图表在正文中必须引用

阅后删除此文本框

位置：表标题在表前面

字体：中文为宋体：西文和数字等为Times New Roman

字号：五号

段落：居中

注意：“表X．X”后留有两个空格

阅后删除此文本框

……

表1.1 软件可靠性研究的发展历程

注意：（1）不能从它处以截图方式获取表格；（2）表需显示所有边框线；（3）图或表均不能独立成节，所有图和表在正文中必须有适量的说明或阐述性文字：（4）表居中

表内文字的字号：五号

字体：中文为宋体：西文和数字等为Times New Roman

阅后删除此文本框

|  |  |
| --- | --- |
| 时间 | 软件可靠性研究发展状况 |
| 1960年以前 | 软件开发过程还处于原始时期，尚没有软件可靠性的概念 |
| 1960-1970年 | “软件危机”出现，人们开始关心软件可靠性问题 |
| 1971-1980年 | 软件可靠性模型的研究成果丰富，软件失效数据的积累、分析工作有了初步的发展 |
| 1981-1990年 | 软件可靠性设计中广泛采用模块化设计、结构化设计方法；复杂系统中己开始采用软件避错技术和容错技术；故障树分析法、故障模式影响分析法也被引入到软件可靠性分析中  正文（含参考文献和附录）页码标注格式为： 1、2、3……  正文页码从1开始  阅后删除此文本框 |
| 1991年至今 | 软件体系结构发生了巨变，传统的软件可靠性理论已明显不能解决现代大规模复杂软件的可靠性问题，软件可靠性的研究又面临着新的挑战 |

1.2主要研究工作

本文的主要研究工作是结合分布式系统及软件可靠性技术的优点，实现增值业务系统的分布式系统设计，以及对软件可靠性的研究，提出了基于增值业务计费的分布式系统的可靠性设计，并加以实现。研究的具体工作如下：

正文与图表之间段间距为0.5行

阅后删除此文本框

……

2 相关技术

2.1分布式系统

注意：文中必须标明每篇参考文献的引用

字体：Times New Roman

字号：小四，上标

阅后删除此文本框

2.1.1 概述

随着网络应用的发展，分布式软件系统得到越来越多的应用。分布式系统多种多样并涉及到不同的系统体系结构，对一些用户来说，一个分布式系统是为了解决单个问题而紧密结合在一起工作的多处理的集合；对另外一些用户来说，一个分布式系统可能意味着一个由地理上分散的处理机组成的计算机网，这些处理机连接在一起以实现不同资源的共享。

三级标题

基准样式：标题3

字体：中文为宋体：西文和数字等为Times New Roman

字号：小四

段落：左对齐

首行缩进2个字符

阅后删除此文本框

对于分布式系统，文献[1]的定义是：“一个分布式系统是一些独立的计算机的集合，但是对这个系统的用户来说，系统就像一台计算机一样。” [2-3]

2.1.2 分布式系统的特点

⑴ 固有的分布式应用

多个连续参考文献引用示例

字体：Times New Roman

字号：小四，上标

阅后删除此文本框

许多应用是固有分布式的，这些应用是突发模式(burst mode)而非批量模式(bulk mode)。这方面的实例有事务处理和Internet Java小程序。这些应用的性能取决于吞吐量(事务响应时间或每秒完成的事务数)而不是应用所用的执行时间

字体、字号等与正文同。阅后删除此文本框

⑵ 性能/价格

分布式系统的并行性降低了处理的瓶颈，提供了全面改进的性能，也就是说，分布式系统提供了更好的性能价格比。……

①模型假设应简单易解，合乎直觉；

②所需数据应便于收集；

③参数估计方法简单可行；

④模型给出结果便于工程确认。

图与正文为0.5行段间距

阅后删除此文本框

⑶ 资源共享

分布式系统能有效地支持不同位置的用户对信息和资源(硬件和软件)的共享。

注意：每个图表在正文中必须引用

阅后删除此文本框

……

如图2.1所示。

注意：（1）必须保证图表清晰美观；图制作要清晰，不能随意缩放，图中和表中文字为五号字。除实现系统的运行界面允许截图外，其余图件原则上须采用Visio绘制。尤其注意，Visio图要采用复制/粘贴的方式嵌入至Word中，不能使用截图方式；（2）图居中

字号：五号字

字体：中文为宋体；西文和数字等为Times New Roman

阅后删除此文本框

NIC

NIC

Host

BUS

Adapter

Local Storage

主机

NIC

NIC

Host

BUS

Adapter

Local Storage

备份机

连向客户端网络

Intra-Cluster LAN

SCSI BUS or Fibred Channel

共享磁盘阵列

图2.1 双机并联系统结构图

注意：图和表要按章编号，例如：第2章的第三个图编号为图2.3，即使该图是第2.4小节的第一个图，也不能编号成2.4.1

字号：五号字

字体：中文为宋体；西文和数字等为Times New Roman

段落：居中

位置：图标题在图后面

正文与图表之间段落设置：段前0.5行，段后0行

阅后删除此文本框

联机相互健康检查是在系统服务的过程中，校验程序每隔一定时间或利用中央处理机的空闲时间自动启动，在两台计算机（计算机A及计算机B）之间相互进行。

# 3 分布式移动

4 结论

注意：正文中在每页的下缘，不能出现大片的空白区域。所有的空白区域要通过前移后文中的文字补上

阅后删除此文本框

参考文献

[1] 罗建幸, 徐红燕. 低市场集中度行业中的市场机会[J]. 商业研究, 2003, 16(276): 138-138.

注意：编号[ ] 与作者间加一个空格

阅后删除此文本框

一级标题，居中

阅后删除此文本框

[2] 徐齐利.互联网平台电商企业的垄断势力与结构——对百度搜索平台客户集中度的测算[J]. 当代财经, 2017, (3): 91-105.

[3] 王领, 刘融. 基于演化博弈论视角的互联网巨头并购研究[J]. 电子商务, 2016, (9): 13-15.

[4] 张新红, 张贝. 2016年中国分享经济发展报告[R]. 北京: 国家信息中心信息化研究部, 中国互联网协会分享经济工作委员会, 2016.

[5] 程维, 柳青. 滴滴: 分享经济改变中国[M]. 北京: 人民邮电出版社, 2016.

[6] Shabtay D., Gaspar N. and Yedidsion L. A bicriteria approach to scheduling a single machine with job rejection and positional penalties[J]. Journal of Combinatorial Optimization, 2012, 23(4): 395-424.

[7] Pinedo M. Scheduling: theory, algorithms, and systems (4th edition) [M]. New York: Springer, 2012.

注意：（1）参考文献不少于15篇；（2）参考文献以正文中引用次序排列；（3）若标明引用页码，必须标明起始两个页码；（4）期刊论文必须标明页码；（5）必须标明文献类型；毕业论文不能少于2篇英文参考文献

中文字体：宋体

字号：五号

英文字体：Times New Roman

字号：五号

段落：左对齐，固定行距20磅(pt)

阅后删除此文本框

致 谢

在本文完成之际，谨向我的导师郭艺辉老师致以衷心的感谢，本论文是在她的精心指导和关怀下完成的，从论文的选题、方案设计，到论文的撰写和修改，都倾注了郭老师的心血和汗水，在学习期间，她的言传身教将使我终生受益，她认真严谨的治学态度、豁达宽广的胸怀、平易近人的处事风格是我一生的楷模，值此提交论文之时，在此向郭老师表达衷心的感谢！

一级标题，居中

阅后删除此文本框

附录 内容名称

以下内容可放在附录之内：

注意：（1）附录与附录内容之间留有一个空格；（2）附录为可选项，根据实际需要确定，若无，将此页删除

一级标题

阅后删除此文本框

(1) 正文内过于冗长的公式推导；

(2) 方便他人阅读所需的辅助性数学工具或表格；

(3) 重复性数据和图表；

(4) 论文使用的主要符号的意义和单位；

(5) 程序说明和程序全文。

字体：宋体

字号：小四

段落：两端对齐，首行缩进2个字符，1.5倍行距

阅后删除此文本框