摘 要

本文基于微服务、大数据分析和机器学习，将此三大热门技术在医疗系统服务中结合并实现。

首先，阐述微服务、大数据分析和机器学习在当今“互联网+医疗”行业的发展现状。

其次，从系统技术架构的角度，对本服务系统进行深入分析并实现业务功能。

第三，结合本服务系统中应用到的微服务、大数据分析组件进行详细分析，以及应用到的机器学习算法原理进行深入说明，并对本服务管理平台的数据处理流程，包括数据采集，数据处理，数据预测，数据反馈等核心步骤进行技术剖析， 最终得出本服务管理平台对社会的价值结论。

最后，对“互联网+医疗”行业的发展做出深入分析。

**关键词：**互联网+医疗；微服务；大数据分析；机器学习；

ABSTRACT

Based on the extensive application of three popular technologies in today's mobile Internet industry: micro services, big data analysis and machine learning, the three popular technologies are combined and realized in this service system.

In this paper, the application of these three popular technologies in the service system will be analyzed in depth.

First, the development of micro services, big data analysis and machine learning in today's mobile Internet industry is expounded.

Secondly, from the perspective of microservices application and architecture, the overall architecture design of the service system is expounded.

Thirdly, a brief introduction and principle analysis of the microservices and big data analysis components applied in the service system, as well as an in-depth explanation of the principles of the machine learning algorithm applied. In addition, the data processing process of the service management platform, including data collection, data processing, data prediction, data feedback and other core steps, is analyzed technically, and finally the conclusion of the value of the service management platform to the society is drawn.

Finally, it makes an in-depth analysis of the development of the mobile Internet industry.

**Keywords**: Mobile Internet; Micro service; Big data analysis; Machine learning; The data analysis

目 录

[摘 要 1](#_Toc524112354)

[ABSTRACT 2](#_Toc524112355)

[目 录 3](#_Toc524112356)

[前 言 4](#_Toc524112357)

[**1** **“互联网+医疗”行业发展现状** 5](#_Toc524112358)

[**2** **系统技术架构设计与分析** 7](#_Toc524112359)

[**2.1** **系统技术架构设计与层次** 7](#_Toc524112360)

[**2.2** **系统技术架构层次阐述** 7](#_Toc524112361)

[2.2.1 网络层 7](#_Toc524112362)

[2.2.2 缓存层 8](#_Toc524112363)

[2.2.3 应用层 9](#_Toc524112364)

[2.2.4 数据计算层 9](#_Toc524112365)

[2.2.5 算法层 10](#_Toc524112366)

[2.2.6 数据存储层 – MySQL 10](#_Toc524112367)

[**3** **核心技术阐述与应用** 11](#_Toc524112368)

[**3.1** **数据采集和数据反馈 - 微服务应用** 11](#_Toc524112369)

[3.1.1 技术 spring-cloud spring-boot 11](#_Toc524112370)

[3.1.2 微服务优点 12](#_Toc524112371)

[3.1.3 微服务技术应用 12](#_Toc524112372)

[**3.2** **数据清洗** 13](#_Toc524112373)

[3.2.1 Kafka 13](#_Toc524112374)

[3.2.2 Storm 14](#_Toc524112375)

[**3.3** **数据计算 - 机器学习应用** 14](#_Toc524112376)

[**4** **发展前景** 15](#_Toc524112377)

[结 论 16](#_Toc524112378)

[致 谢 17](#_Toc524112379)

[参考文献 18](#_Toc524112380)

[附 录 19](#_Toc524112381)

前 言

在当今的互联网时代，微服务、大数据分析处理和机器学习的广泛应用已经深入到所有行业，所以，互联网医疗，作为互联网在医疗行业的新应用，代表了医疗行业新的发展方向，有利于解决中国医疗资源不平衡和人们日益增加的健康医疗需求之间的矛盾，是国家积极引导和支持的医疗发展模式。

近年来，随着国家放开二胎政策，生宝宝的妈妈越来越多，全国大大小小的妇幼保健院的办公压力也逐年增加。由于这些妇幼保健院多数是早期建造的，在网络硬件设施上，承受着很大负荷。所以，移动智能办公在当今互联网时代，体现着举足轻重的作用，凭借其智能、方便、易操作、反应快速、承载信息量大等多方面的优势，越来越受人们的青睐。同时，微服务技术的飞速发展日趋完善，很大程序推动了移动智能办工的使用，在互联网时代，以及刚刚产生的物联网都得到广泛应用。

本课题以准妈妈孕期信息服务管理作为平台，搭载智能手机、智能平板等移动终端设备，实现对准妈妈孕期的身体状态、孕检信息等数据的采集，在PC端通过服务平台，实现对采集数据的预测，达到对准妈妈状况的实时监控和各种预测的目的。

本服务系统可基于多种类型的硬件设备及不同的操作系统实现部署的多样化，并基于大数据组件的高兼容，高并发，高性能的处理能力，提高数据采集、数据清洗、数据计算和数据反馈等核心技术能力，以及利用机器学习算法，实现智能预测及信息的准确推送。

# **“互联网+医疗”行业发展现状**

2018年，沉寂两年的互联网医疗重回头条。

四年前，互联网医疗迎来了大热元年，但很快就遭遇资本寒冬。熬过生死线后，幸存的纯线上问诊难有盈利点，转型迫在眉睫。至于那些转换赛道死磕线下服务的项目，也很难说从轻问诊转型重服务何时才能奏效。

但一系列的政府动作正在释放利好信息。4月11日，李克强总理视察上海华山医院，称赞互联网远程医疗。一天之后，国务院常务会议通过了“互联网+医疗健康发展意见”。5天后国新办又公开介绍了相关“发展意见”精神。当月26日，卫健委发布“互联网+医疗健康”实施计划。

在互联网医疗创业者看来，这意味着“互联网+医疗”首度从国家战略的层面被认可。

从2014年“移动医疗创业元年”，到2016年开始的资本寒冬，“互联网+医疗”在短短两年内经历了过山车般的大起大落。

首先，让我们先了解一下，互联网医疗的主要内容？目前我国具代表性的互联网医疗形式？

互联网医疗，作为依托于互联网技术在医疗行业的新应用尝试，包括了以互联网为载体和技术手段的健康教育、医疗信息查询、电子健康档案、疾病风险评估、在线疾病咨询、电子处方、远程会诊、及远程治疗和康复等多种形式的健康管家服务。

近几年在国内有了一定程度的发展，具有代表性的互联网医疗形式有：

1、以健康教育和信息为主的39健康网；

2、以患者社区和医生信息为主的爱好医生；

3、以医师评价和挂号为主的好大夫在线；

4、以电子健康档案采集和应用为主的120ehr网；

5、以疾病风险评估为主的宜康网；

6、以即时在线咨询为主的医通无忧网等；

7、以远程云诊、全程陪诊为主的[就诊通](https://baike.baidu.com/item/%E5%B0%B1%E8%AF%8A%E9%80%9A)网；

8、以社区医院，患者签约家庭医生，医院沟通为主的身边医生（app）；

9、以专科垂直领域为主，专注[眼科](https://baike.baidu.com/item/%E7%9C%BC%E7%A7%91/38035)服务的[目邻](https://baike.baidu.com/item/%E7%9B%AE%E9%82%BB)；

在互联网企业开始关注面向医院的全流程服务，全景医疗的概念开始出现，同时，以提升就医体验为目的的医疗服务o2o模式备受关注，越来越多的企业开始关注对医疗活动各个阶段中所产生的数据进行采集、存储和处理，医疗大数据被提到重要位置。随着医疗大数据应用的进一步推进，我国当前医疗资源配置碎片化导致的数据碎片化已经成为政府和产业都高度重视并着力解决的问题，由此引出了产业对电子病历、医疗影像数据化、临床数据、生物医药数据等领域的关注。

因此，本服务系统志在利用“互联网+医疗”这个新兴的医疗应用，实现并帮助准妈妈在孕期过程中，了解更多的知识，简化各个阶段的检查步骤，利用机器学习方法实现对准妈妈怀孕期间健康情况的指导建议，并实现碎片化的医疗数据的整合。

# **系统技术架构设计与分析**

## **系统技术架构设计与层次**



通过上图所示，本服务系统基于6层架构设计并实现，从总体来说，这样的架构设计会使服务系统层次清晰，降低各层服务的耦合度，层与层服务的开发互不干扰，实现起来更为直接简单快捷，更能适应当前互联网行业的架构设计理念。

## **系统技术架构层次阐述**

### 网络层

在传统的Web项目中，并发量小，用户使用的少。所以在低并发的情况下，用户可以直接访问tomcat服务器，然后tomcat服务器返回消息给用户。

但为了解决高并发，通常情况下，我们会使用负载均衡：也就是我们多增加几个WEB服务器。当用户访问的时候，请求可以提交到空闲的tomcat服务器上。当网站的访问量达到一定程度后，单台服务器不能满足用户的请求时，需要用多台服务器集群可以使用nginx做反向代理。并且多台服务器可以平均分担负载，不会因为某台服务器负载高宕机而某台服务器闲置的情况。

### 缓存层

在本服务系统中，使用Redis作为数据缓存层的存储工具，其使用的目的：

1、将热点数据，例如：用户登录信息等，放入缓存数据是为了让客户端很少或者甚至不访问数据库服务器进行的数据查询。在高并发下，最大程度降低对数据库服务器的访问压力。

2、防数据穿透，对缓存查询加锁，如果KEY不存在，就加锁，然后查DB入缓存，然后解锁；其他进程如果发现有锁就等待，然后等解锁后返回数据或者进入DB查询。

那么，现在缓存数据库种类这么多，为什么使用Redis呢？

使用Redis不仅可以让代码变得更简短、更易懂、更易维护，而且还可以使代码的运行速度更快（因为用户不需要通过读取数据库来更新数据）。除此之外，在其他许多情况下，Redis的效率和易用性也比关系数据库要好得多。

**1、内存方面：**Redis将键值存储在主存中，用于快速地读写访问。

**2、**Redis支持主从复制。数据读取在slave完成，而数据写入在 master 完成。复制提供可伸缩性和可用性。任何一个slave宕机，其他的slave还可以提供数据访问。

**3、数据结构**：Redis不仅存储字符串，还支持列表，集合，哈希和有序集合。Redis的聚合数据可以是整数或者浮点数，而memcached的聚合数据只能是整数。

**4、虚拟内存**：Redis使用RAM作为内存式存储。但是，在内存不足的情况下，它使用虚拟内存来保存数据。

**5、发布 / 订阅模型**：Redis支持创建发布和订阅通道，这样Redis客户端可以订阅任意的通道来进行数据消费，并且任何已订阅该通道的客户端可以发布数据。

6、**数据持久性**：Redis将内存中的数据定期保存到文件系统中。当Redis节点故障时，数据可以从Redis数据文件恢复。

### 应用层



在本服务系统中，应用层使用Spring Boot框架，并设置：用户管理，信息管理、数据管理、系统管理和日志管理模块。

用户管理模块：维护着准妈妈和医师信息，并将热点数据如准妈妈和医师的登录信息等放置于缓存数据库中。

信息管理模块：提供准妈妈和医师信息的更新和修改功能，以及准妈妈检查结果信息和孕检过程等流程信息。

数据管理模块：提供多维度的数据统计展示。

系统管理模块：提供孕检预约功能、信息管理功能和检查报告进度展示功能。

日志管理模块：收集各个功能模块执行过程的日志，并提供告警和多维度统计功能。

### 数据计算层

#### 实时流程



在本服务系统的实时，使用Kafka和Storm作为数据计算层的服务组件。

在Kafka中，通过创建用户信息Topic、检查预约Topic、检查告警Topic和用户反馈Topic，WEB应用可根据用户业务请求将相应的请求数据放置于相应的TOPIC队列中。

在Storm中，定制相应的业务获取相应Topic队列中数据，依次进行数据清洗、格式转换、业务处理、生成日志和数据入库等功能。

数据清洗用于处理无效的请求数据或垃圾数据，减少占用不必要的数据处理资源。

格式转换用于将用JSON格式定义的请求数据转换成MySQL和Elastic Search中定义的数据结构。

业务处理即为针对应用层定制相应的业务数据的处理流程。

生成日志即针对整个数据清洗层各个模块的处理功能，生成相应的处理日志，并进行实时监控，对发生异常的情况进行实时提醒。

数据入库即当正常数据处理完成后，将有结果数据写入到最终的数据存储中。在本服务系统中，将写入到MySQL和Elastic Search。

#### 离线流程



在本服务系统的离线流程中，包含了两部分处理流程：

1. 数据回溯即实时流程的全部处理过程。当服务系统发生异常，不能正常处理，造成Kafka中若干Topic队列出现积压，致使产出的最终数据不正确。在这种情况下，我们就必须将一段时间内的原始数据通过离线流程进行数据回溯，把最终产出的数据进行数据补全，并重新写入到数据库中。而对于已经存在的数据，将进行数据更新，以供Web应用层进行正确的数据检索。
2. 算法模型数据迭代，会针对不同的数据如准妈妈每次检查报告数据和身体、饮食状况的数据分别进行相应业务数据挖掘，挖掘出更多报告中隐含的数据信息，并由营养师结合全部数据信息给出更加专业地指导建议。

### 算法层

算法层是整个数据架构的关键点，所有底层数据发挥作用都需要经过数据挖掘和计算框架输出。数据计算层既要通过不同算法满足不同需求的挖掘需要，又要根据数据需求尽可能的实时输出结果。

### 数据存储层

在本服务系统中，采用MySQL作为数据的最终存储、Elastic Search作为Web应用层进行检索的数据存储。

#### MySQL



在本服务系统中，从内容上来说，会创建准妈妈的相关数据表，如准妈妈基础信息表、预约信息表、日常身体饮食情况记录表和孕检结果信息表，及医师基础信息表等。

准妈妈基础信息表存储在医院建档的准妈妈信息，如姓名、身份证号、籍贯、年龄、最后例假等信息。

医师基础信息表存储医院在册的医师的详细，准妈妈可根据自己的意向选择合适自己的医师，进行之后的每次孕检。

预约信息表存储准妈妈预约下次孕检的时间与医师的信息。当准妈妈预约成功后，到当日检查，本服务系统以短信方式进行提醒，以免过期，错过检查。

日常身体饮食情况记录表存储准妈妈每天的身体和饮食相应信息，这部分数据用于数据挖掘和预测业务，从中挖掘出更多的信息提供给准妈妈和准妈妈指定的医师。

孕检结果信息表存储准妈妈每次孕检结果报告的数据，及经过数据挖掘后的数据信息，并结合医师和营养师的建议，得出更为详细的数据报告。

采用MySQL作为数据的最终存储，是由于其具有如下特点：

1. 它使用的核心线程是完全多线程，支持多处理器。
2. 有多种列类型：1、2、3、4、和8字节长度自有符号／无符号整数、FLOAT、DOUBLE、CHAR、VARCHAR、TEXT、BLOB、DATE、TIME、DATETIME、 TIMESTAMP、YEAR、和ENUM类型。
3. 它通过一个高度优化的类库实现SQL函数库并像他们能达到的一样快速，通常在查询初始化后不该有任何内存分配。没有内存漏洞。
4. 全面支持SQL的GROUP BY和ORDER BY子句，支持聚合函数(COUNT()、COUNT(DISTINCT)、AVG()、STD()、SUM()、MAX()和MIN())。你可以在同一查询中混来自不同数据库的表。
5. 支持ANSI SQL的LEFT 0UTER JOIN和ODBC。
6. 所有列都有缺省值。你可以用INSERT插入一个表列的子集，那些没用明确给定值的列设置为他们的决省值。
7. MySQL可以工作在不同的平台上。支持C、C++、Java、Perl、PHP、Python和TCL API。

#### Elastic Search

采用 Elastic Search 作为Web应用层的检索存储，是由于其具有如下特点：

1. 分布式实时文件存储，可将每一个字段存入索引，使其可以被检索到。
2. 实时分析的分布式搜索引擎。分布式：索引分拆成多个分片，每个分片可有零个或多个副本。集群中的每个数据节点都可承载一个或多个分片，并且协调和处理各种操作；负载再平衡和路由在大多数情况下自动完成。
3. 可以扩展到上百台服务器，处理PB级别的结构化或非结构化数据。
4. 支持插件机制，分词插件、同步插件、Hadoop插件、可视化插件等。

# **核心技术阐述与应用**

## **数据采集和数据反馈 - 微服务应用**

### 技术 spring-cloud spring-boot

微服务架构是一种架构模式，它提倡将单一应用程序划分成一组小的服务，服务之间互相协调、互相配合 ，为用户提供最终价值。每个服务运行在其独立的进程中，服务于服务间采用轻量级的通信机制互相沟通（通常是基于HTTP的RESTful API）。每个服务都围绕着具体业务进行构建，并且能够被独立地部署到生产环境、类生产环境等。另外，应尽量避免统一的、集中式的服务管理机制，对具体的一个服务而言，应根据业务上下文，选择合适的语言、工具对其进行构建。

—摘自 马丁·福勒先生的博客

通常情况下，我们会将Spring Cloud, Spring Boot 作为微服务的架构来使用。所以，在本服务系统中，我将使用Spring Boot来进行开发。

载体：各种客户端：手机，IPAD，手环装入APP

### 微服务优点

首先，通过分解巨大单体式应用为多个服务方法解决了复杂性问题。在功能不变的情况下，应用被分解为多个可管理的分支或服务。每个服务都有一个用RPC-或者消息驱动API定义清楚的边界。微服务架构模式给采用单体式编码方式很难实现的功能提供了模块化的解决方案，由此，单个服务很容易开发、理解和维护。

第二，这种架构使得每个服务都可以有专门开发团队来开发。开发者可以自由选择开发技术，提供API服务。当然，许多公司试图避免混乱，只提供某些技术选择。然后，这种自由意味着开发者不需要被迫使用某项目开始时采用的过时技术，他们可以选择现在的技术。甚至于，因为服务都是相对简单，即使用现在技术重写以前代码也不是很困难的事情。

第三，微服务架构模式是每个微服务独立的部署。开发者不再需要协调其它服务部署对本服务的影响。这种改变可以加快部署速度。UI团队可以采用AB测试，快速的部署变化。微服务架构模式使得持续化部署成为可能。

最后，微服务架构模式使得每个服务独立扩展。你可以根据每个服务的规模来部署满足需求的规模。甚至于，你可以使用更适合于服务资源需求的硬件。

### 微服务技术应用

由于当今互联网WEB开发，大部分基于Spring框架，并以友好的用户界面为用户提供所需的各项应用软件和服务。其好处在于：

1. 方便解耦，简化开发

Spring就是一个大工厂，可以将所有对象创建和依赖关系维护，交给Spring管理。

1. AOP编程的支持

Spring提供面向切面编程，可以方便的实现对程序进行权限拦截、运行监控等功能。

1. 声明式事务的支持

只需要通过配置就可以完成对事务的管理，而无需手动编程。

1. 方便程序的测试

Spring对Junit4支持，可以通过注解方便的测试Spring程序。

1. 方便集成各种优秀框架

Spring不排斥各种优秀的开源框架，其内部提供了对各种优秀框架（如：Struts2、Hibernate、MyBatis、Quartz等）的直接支持。

1. 降低JavaEE API的使用难度

Spring 对JavaEE开发中非常难用的一些API（JDBC、JavaMail、远程调用等），都提供了封装，使这些API应用难度大大降低。

## **数据清洗**

### **Kafka**

#### 简述

**Kafka**是由Apache软件基金会开发的一个开源流处理平台，由Scala和Java编写。Kafka是一种高吞吐量的分布式发布订阅消息系统，它可以处理消费者规模的网站中的所有动作流数据。

#### 特点

Kafka[1]  是一种高吞吐量[2]  的分布式发布订阅消息系统，有如下特性：

* 1. 通过O(1)的磁盘数据结构提供消息的持久化，这种结构对于即使数以TB的消息存储也能够保持长时间的稳定性能。
  2. 高吞吐量[2]  ：即使是非常普通的硬件Kafka也可以支持每秒数百万[2]  的消息。
  3. 支持通过Kafka服务器和消费机集群来分区消息。
  4. 支持Hadoop并行数据加载。[3]
  5. 可扩展性：kafka集群支持热扩展。
  6. 容错性：允许集群中节点失败（若副本数量为n,则允许n-1个节点失败）

参考：

1.Apache Kafka官网  ．Apache Kafka官网[引用日期2018-1-5]

2.Benchmarking Apache Kafka: 2 Million Writes Per Second (On Three Cheap Machines)  ．LinkedIn．2014-04-27[引用日期2015-11-20]

3.分布式发布订阅消息系统 Kafka  ．开源中国社区．2011-05-07[引用日期2015-06-29]

### **Storm**

#### 简述

随着大数据业务的快速增长，针对大规模数据处理的实时计算变成了一种业务上的需求，缺少“实时的Hadoop 系统”已经成为整个大数据生态系统中的一个巨大缺失。Storm 正是在这样的需求背景下出现的，Storm 很好地满足了这一需求。

在Storm 出现之前，对于需要实现计算的任务，开发者需要手动维护一个消息队列和消息处理者所组成的实时处理网络，消息处理者从消息队列中取出消息进行处理，然后更新数据库，发送消息给其他队列。所有这些操作都需要开发者自己实现。这种编程实现的模式存在以下缺陷。

#### 特点

在使用过程中，其优点主要有以下几点：

1. 简单的API接口
2. 可扩展性
3. 容错性
4. 无数据丢失
5. 支持多种编程语言
6. 容易部署和操作

## **数据计算 - 机器学习应用**

# **发展前景**

提高效率，信息化，减少人员消耗

结 论

本论文所取得的成果

在这次毕业设计里，深入地接触微服务架构的搭建、大数据平台环境的搭建和机器学习的深入学习。最终，设计并制件出本服务系统，体验到移动互联网技术应用于医疗行业，以至于应用到其他行业所带给我们的便利。

“互联网+医疗”的发展浅析

就医难是国内医疗面临的最大问题，以互联网化为手段，优化就诊流程，提升患者的就医体验，将会成为未来“互联网+医疗”一直关注的问题。因此，一切围绕这个目的的服务会一直存在，可能贯穿医疗服务的全过程，具体可涵盖：医疗资源查找与匹配、网上挂号、在线问诊、远程诊疗、医药电商、移动医疗等领域。

互联网环境下，医疗服务逐步数字化，这将极大地提高医生与患者直接的相互了解，例如透过先进的影像获取和存储技术，利用大数据分析，获得特定个人的病灶变化情况及同类病例治疗的比较结果，这将为精准治疗提供决策依据。将患者的医疗服务需求精准推送给医生，将医疗服务项目精准提供给患者，实现医疗沟通过程中的双向精准化，才能真正达成精准治疗。

在“互联网+医疗”的驱动下，医疗机构将建立起以患者为中心的全新医疗服务模式，以改善就医体验为目的，逐步实现医疗诊治精准化、医疗组织协同化、医疗服务个性化，将医疗服务扩展到更大范围。未来医疗新模式将在信息技术的推动下向共享、协作、个性化方向发展，“互联网+医疗”将会涌现出更多的应用和模式，而这需要进一步的研究和探索。

致 谢

……

参考文献

[1] 张思莱.张思莱育儿手记.中国妇女出版社.2017年5月第6次出版

[2] 李航. 统计学习方法. 清华大学出版社. 2012年3月第1次出版

[3] 周志华. 机器学习. 清华大学出版社. 2016年1月第1次出版

[4] 罗聪翼, 龚成志译. Storm应用实践：实时事务处理之策略. 机械工业出版社. 2018年1月第1次出版

[5] Neha Narkhede, Gwen Shapira, Todd Palino. Kafka: The Definitive Guide.

[6] 朱松岭. 离线和实时大数据开发实战. 机械工业出版社. 2018年5月第一次出版

[7] Nick Mcclure. TensorFlow Machine Learning Cookbook. 2017.

[8] John Carnell. Spring Microservices IN Action. 2018.

[9] 郑天民. 微服务设计原理与架构. 人民邮电出版社. 2018年5月第1次出版

[10] [美] Sanjay Patni 著 郭理勇 译. RESTful API开发实战. 清华大学出版社. 2018年2月第1次出版

[11] 妇幼医院相关孕检说明

附 录

……