|  |  |
| --- | --- |
| **民大校徽** | **中央民族大学课程设计报告** |
| **Course Lab Report in MUC** |

**新冠疫情分析系统**

**需求分析说明书**

|  |  |
| --- | --- |
| 小组成员： | 20011441 黄影 |
| 小组成员： | 20150019 李承骏 |
| 院 系 ： | 信息工程学院 |
| 专 业： | 软件工程 |
| 任课教师： | 孙娜 |

目录

[1、 文档介绍 4](#_Toc7745)

[1.1文档目的 4](#_Toc2235)

[1.2文档范围 4](#_Toc3571)

[1.3读者对象 4](#_Toc3144)

[2、 系统介绍 4](#_Toc4991)

[2.1系统概述及用途 4](#_Toc12961)

[2.2开发背景 5](#_Toc17557)

[3、 系统功能性需求 5](#_Toc22430)

[4、 系统非功能性需求 6](#_Toc16132)

[5、 面向的用户团体 7](#_Toc30144)

1. **文档介绍**

**1.1文档目的**

本文档旨在介绍基于Spark的新冠疫情分析系统的设计与实现，利用新冠肺炎疫情监测平台，将全国疫情信息、基本物资、病例信息等信息进行即时采集，更能直观地掌握疫情形势,利用 Java进行数据收集、数据处理、数据分析、数据处理、数据分析等。

**1.2文档范围**

本文档包含以下几个部分：

(1) 系统的设计概念；

(2) 系统的可行性分析；

(3) 系统功能模块设计；

(4) 数据模型设计；

(5) 应用原型设计；

**1.3读者对象**

本文档面向多种读者对象：

1. 项目经理：项目经理可以根据该文档了解预期系统的功能。
2. 设计员：对需求进行分析，并完成数据库的设计。
3. 程序员：了解系统功能，据此编写《用户手册》。
4. 测试员：根据本文档编写测试用例，并对软件产品进行功能性测试和非功能性测试。
5. 用户：了解预期产品的功能和性能，并与分析人员一起对整个需求进行讨论和协商。在阅读本文档时，首先要了解产品的功能概貌，然后可以根据自身的需要对每一功能进行适当的了解。
6. **系统介绍**

**2.1系统概述及用途**

本系统使用网络爬虫爬取疫情数据，在项目中创建crawler模块，专门用于负责数据的爬取部分并使用正则表达式获取 json 格式的疫情数据。通过JSON.toJSONString()方法将JSON格式的数据转换成字符串类型，方便我们将生成的疫情物资数据转换为jsonStr再发送到Kafka集群来进行数据分析。采用爬虫、数据库、大数据分析、爬虫理解、使用、数据库（MySQL）和大数据知识等方法进行。本系统的用途有：

1.有效支持科学判断。大数据报告能将大量的数据信息以生动形象的展示方式呈现在观众面前，在经过整体的数据汇总、分析后，大众能快速从大数据展示中，得出初步的判断，从而加速决策者的处理效率。

2.信息的有效传递。数据信息所面向的不仅是决策者，也能向大众进行授权展示，我们可以通过疫情实时大数据报告了解到全国各地、世界各地的感染人数，以此了解疫情的变化。

3.加速指令执行。大数据报告将各项信息进行汇总、处理，以形象的展示形式，呈现给大众，在此之中，能加速用户指令的执行，实现高效、快速的应用。

**2.1.1疫情分析的需求**

自2023年1月8日起，我国解除对新型冠状病毒感染采取的《中华人民共和国传染病防治法》规定的甲类传染病预防、控制措施 ,为了让疫情的累积确诊和当天确诊、累积治愈和当天治愈、累积死亡和当天死亡等数据更直观的呈现，通过对疫情数据的爬取转换，大数据分析处理数据将疫情情况形象地展示给受众。

**2.2开发背景**

随着变异株毒力减弱、疫苗接种普及、防控经验积累，中国宣布从2023年1月8日起优化调整疫情防控政策，取消大部分防疫限制措施，中国公民跨境旅行也将恢复。在这种情况下，对疫情的数据监管更加必不可少，疫情数据的分析对新冠病毒的控制和预防具有重要意义。

1. **系统功能性需求**

### **3.1爬取疫情数据功能分析**

在项目中创建crawler模块，专门用于负责数据的爬取部分。通过HttpUtils.getHtml()方法来爬取指定页面并解析页面中的指定内容。使用正则表达式获取 json 格式的疫情数据

### **3.2物资信息生产分析**

通过JSON.toJSONString()方法可以将JSON格式的数据转换成字符串类型，方便我们将生成的疫情物资数据转换为jsonStr再发送到Kafka集群来进行数据分析。

**3.3数据处理分析**

创建StructuredStreaming执行环境，从kafka接收消息并取出消息中的value，输出目的地，输出模式,默认就是append表示显示新增行；

**3.4数据存储分析**

首先应建立基本信息材料数据库，以存储有关材料数量及其变化的信息。数据存储器可以记录在获取新数据时发生的变化，并使我们能够对数据进行更多的分析，同时也可以更安全地长期存储这些变化。

1. **系统非功能性需求**

### **4.1系统安全性**

考虑到每一个系统的运作都是稳定、稳定和稳定的，不能出错，运作过程亦不容易崩溃，因此，必须在不同程度上为系统工程的策划制订明确的规管和制度规管，详情如下：

大数据分析项目开发的系统设计，分析了数据的每一部分，都是事先编写的，因此数据源的安全性和稳定性是整个项目最重要的。如果未经安全验证的数据进入分析程序，可能会导致未知的系统错误，并且Kafka群集被禁用甚至更新。因此，在获取数据时，必须严厉谴责要发送给CAFCA集群的空操作、非法参数验证操作和安全数据验证操作。而征服整个页面的时间也必须进行一点判断，确保数据源的稳定安全，避免项目出错时出现太严重的错误。因此，系统的安全性更高。

### **4.2数据完整性**

如果要求是出于维护数据完整性的需要，则限制条件必须设置在数据库级别。如果在数据层确实引入了约束，那么必须保证有一种方法可以让用户界面知道，由于用户使用用户理解的语言所理解的原因，无法获得更新。除非字段名是多义的，否则字段名本身是不够的。请在任何可能的情况下使用数据库系统来实现数据的完整性。这不仅包括通过标准化实现的完整性，还包括数据的功能性。在编写数据时，还可以添加触发器，以保证数据的真实性。不要依赖商业级的数据完整性保证；它不能保证表（对外关系）的完整性，因此它不能强加在其他完整性规则之上。

对于分布式系统，在决定是否将所有数据复制到不同的站点或存储在一个地方之前，应该估计未来五年或十年的数据量。当您将数据传输到其他站时，最好在数据库字段中放置多个标记。在目标站点收到您的数据后更新您的标记。要执行此传输，请编写您自己的处理批次或控制程序，该程序在一定的时间间隔内运行，不允许用户在日常运行后传输数据。本地复制您的服务数据，如计算常数和利润率，安装版本保证每个站点的数据完全匹配。

没有什么好方法可以在有害数据进入数据库后删除它们，所以必须在它们进入数据库之前将它们截断。激活数据库系统的指示完整性。这将使数据保持干净，并迫使开发人员花费更多的时间处理不正确的条件。

1. **面向的用户团体**

|  |  |
| --- | --- |
| 用户类型 | 作用 |
| 有关部门 | 及时调整政策 |
| 大众 | 及时了解疫情的变化，做好个人防护 |
| 医护人员 | 便于准备对已感染人员的医护工作 |
| 管理员 | 方便管理平台所有数据 |