运营商网络流量流向异常账号监测系统概要设计说明书

目录

[1引言 3](#_Toc23792)

[1.1编写目的 3](#_Toc25038)

[1.2背景 3](#_Toc6552)

[1.3术语解释 3](#_Toc8701)

[2系统总体设计描述 3](#_Toc12465)

[3各个模块设计描述 5](#_Toc10075)

[3.1数据来源模块描述 5](#_Toc19390)

[3.2 数据存储模块描述 5](#_Toc13151)

[3.3数据处理模块描述 5](#_Toc29779)

[3.4数据展示模块描述 5](#_Toc10077)

[4. 数据库设计 5](#_Toc21565)

[5. 流程图 6](#_Toc17558)

[6. 类及方法的定义 6](#_Toc29651)

# 1引言

## 1.1编写目的

利用比较抽象的语言对整个需求进行概括，确定对系统的物理配置，确定整个系统的处理流程和系统的数据结构，接口设计，人机界面，实现对系统的初步设计。

## 1.2背景

随着时代的发展，目前有很多不法分子根据运营商系统的漏洞，通过多账号共享的问题，给运营商利益带来了巨大的挑战。我们就是运用大数据技术，从海量的数据中挖掘出异常的网络流量流向异常账号，并对其限制，从而缓解对运营商利润的冲击。

## 1.3术语解释

|  |  |
| --- | --- |
| 术语 | 解释 |
| Hadoop | 分布式系统的基础架构 |
| Flume | 一个高可用的，高可靠的，分布式的海量日志采集、聚合和传输的系统 |
| Kafka | 一个分布式的，高吞吐，可扩展的消息发布-订阅系统，用于缓存数据 |
| Storm | 一个开源的，分布式，高容错的实时计算系统 |
| 数据可视化 | 用Servlet+Jsp+Ajax+Echarts+MySQL技术，把统计出的数据展示出来 |

# 2系统总体设计描述

根据前面的功能需求分析和系统整体的需求，将整个运营商在线服务系统划分为以下几层：数据来源层、数据存储层、数据处理层和数据展示层。图3-1为从数据来源层到数据展示层的结构图：

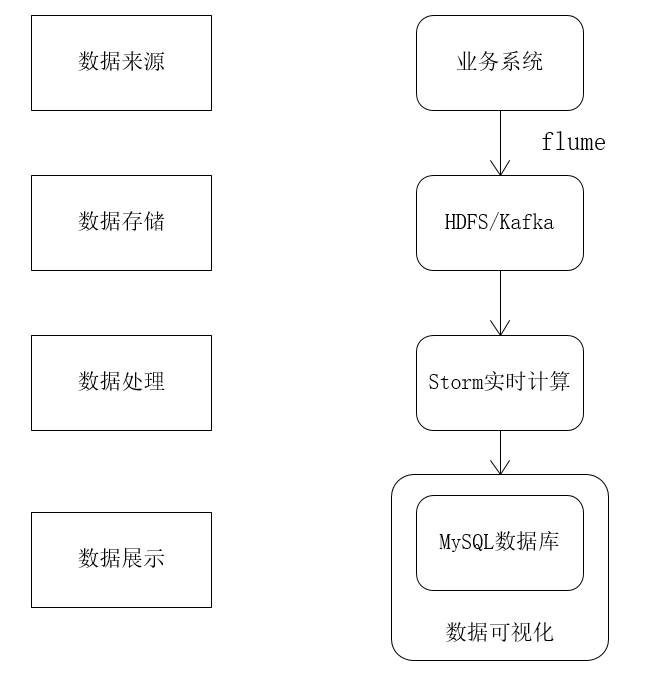


图2-1 运营商网络流量流向异常账号监测系统平台结构图

运营商网络流量流向异常账号监测系统的总体设计就是从运营商那里拿到数据信息进行分析异常的账号信息，设计该系统的目的是从海量数据中检索出异常的账号信息，在详细设计时给出具体的架构。

为了最后实现该系统，在总体设计中把握以下重点，然后在详细设计中可以分别对每个细节难点和重点一一解决。总体设计中包括以下核心的问题，在进行程序详细设计时需要针对这些个难点一一对应的进行分析：

(1)各大运营商每天产生的数据量是非常庞大的，但由于没有数据，需要人工造一些假数据，通过flume采集这些假数据到HDFS或者为了更高的效率，将数据存入kafka缓存中，所以flume组件合理地设计和使用是获取和数据存储的关键。

(2)页面与数据库MySQL接口的对接问题。在页面开发人员对数据进行操作时，Web要与MySQL进行交互，它们两者的接口就需要互相统一，并且要有较高的效率。所以在编写MySQL接口的时候，需要考虑到它们的性能、效率、优化等问题，让接口符合系统的要求。

# 3各个模块设计描述

## 3.1数据来源模块描述

不管一个系统是多么复杂、多么简单，它们的共同点就是都需要符合要求的数据，这样才可以接下来的操作。也可以这么说，如果一个系统代码是实现系统功能的主要部分，那么数据就是核心。因此该系统的数据也是一个值得关注的部分。该项目的数据来源于服务商的业务系统分为两个部分：一个是传统的关系型数据库，一个是每天上传的工单历史记录。

## 3.2 数据存储模块描述

本模块主要有两部分，第一部分是数据源放入到Hadoop的HDFS分布式文件系统中或Kafka缓存消息队列，第二部分就是把处理好的数据存储到MySQL数据库中。

## 3.3数据处理模块描述

该模块用到的技术是Hadoop生态圈中的Storm分布式实时计算框架。运用Storm对HDFS上的文件或者是Kafka缓存中的数据进行处理，最后把处理好的数据存储到MySQL数据库中。

## 3.4数据展示模块描述

通过数据可视化技术将分析结果绘制图表展示到Web页面，为运营商检测异常账号信息提供支持。

# 数据库设计

由于本次项目是用来实时计算的，所以注重的是时效性，由此简单的设计了一下异常表的结构，字段有

id int primary key auto\_increment,

time varchar(50),

userAccount varchar(25),

userIP varchar(15),

qqid varchar(15),

natIP varchar(15),

cookieValue varchar(64),

devName varchar(50),

osName varchar(50)

# 流程图

主要是程序的执行过程，以画图的形式展示。

# 类及方法的定义

介绍每个类是干什么的，具体的方法的作用。