

基于大数据车辆模拟运行及日志收集仿真研究

任培花, 李健浩

(山西大同大学 计算机与网络工程学院, 山西 大同 037009)

摘要: 中国汽车保有量呈井喷式上涨, 需要面对越来越多的车辆违法行为, 如超速行驶、遮挡牌照、套牌、占道等等。针对这些问题, 传统的车辆稽查方式显得力不从心, 存在数据单一、实时性差、离线分析为主等缺点。为减轻交管部门的车辆稽查压力, 构建了以大数据实时分析技术为基础的车辆模拟运行及日志收集仿真模拟系统。该仿真模拟系统采用 Flume、Hadoop 框架和 Redis 集群, 参照城市街道布局设计, 不仅对海量车辆的超高并发度运行和城市交通进行了模拟, 还对车辆日志信息进行了实时收集。以此为依据, 构建的车辆模拟运行场景, 可以有效地提高车辆稽查效率。该系统为数据量较大的行业提供了一种解决问题的思路, 用同样的方法可以为城市旅游、城市产业分布等进行场景模拟, 进而调整旅游服务、产业调整等方案。

关键词: 车辆模拟; 日志收集; 大数据; Flume; Hadoop

中图分类号: U495

文献标识码: A

文章编号: 1673-629X(2020)02-0197-05

doi: 10.3969/j.issn.1673-629X.2020.02.038

Simulation Research on Vehicle Simulated Operation and Log Collection Based on Big Data

REN Pei-hua, LI Jian-hao

(School of Computer and Network Engineering, Shanxi Datong University, Datong 037009, China)

Abstract: As car ownership in China is rising, it is necessary to face more and more illegal vehicles, such as speeding, blocking licenses, fake plate, occupying lanes and so on. In view of these problems, the traditional vehicle inspection method seems to be inadequate, with shortcomings such as single data, poor real-time performance, off-line analysis and so on. In order to alleviate the pressure of vehicle inspection in traffic control department, we construct a simulation system of vehicle simulation operation and log collection based on large data real-time analysis technology. This system uses Flume, Hadoop framework and Redis cluster. Referring to the layout design of urban streets, it not only simulates the super-high concurrent operation of a large number of vehicles and urban traffic, but also collects real-time vehicle log information. Based on this, the vehicle simulation scenario can effectively improve the efficiency of vehicle inspection. This system provides a solution for industries with large data. The same method can be used to simulate scenarios of urban tourism and industrial distribution and then adjust tourism services, industrial adjustment and other programs.

Key words: vehicle simulation; log collection; big data; Flume; Hadoop

0 引言

车辆稽查^[1]是指, 通过道路执法系统对过往车辆信息的实时采集, 并将拍摄日志实时传回交警指挥中心的服务器进行存储, 然后对海量日志数据进行处理, 从中找出嫌疑车辆。

传统的车辆稽查方式应对日益上涨的交通执法压力显得力不从心, 尤其面对海量信息, 采用人工或半人工的稽查处理方式不仅费时费力, 而且滞后严重, 这些

情况说明传统的车辆稽查方式显然是不合理的。因此, 在海量车辆信息和日志信息的大背景下, 如何降低车辆违法行为的稽查压力, 是当前急需解决的问题。

关于车辆稽查系统的研究国内起步较晚, 是随着汽车保有量的增长, 车辆各种违法行为(违法行驶、遮挡拍照、套牌等)增多的现状推动下, 对车辆稽查系统问题的研究才逐渐开始的。很多学者都是借助物联网进行采集, 然后直接将数据保存、后期分析的方法开展

收稿日期: 2019-03-12

修回日期: 2019-07-12

网络出版时间: 2019-09-25

基金项目: 山西省高等学校教学改革创新重点项目(J2017093/2017); 山西省高等学校教学改革项目(2015090/2015); 大同市科技局软科学(2016120/2016)

作者简介: 任培花(1980-), 女, 硕士研究生, 副教授, 研究方向为大数据技术、数据挖掘; 李健浩, 软件工程师, 研究方向为大数据技术。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190925.1525.072.html>

研究。文中减少了对物联网技术的依赖,以稽查数据研究为重点,借助大数据技术^[2],提出一种交通运行数据模拟方案。在具体技术实现方面,在前人研究地基础上,进行了一系列创新。如江日念等提出将 Maven 应用到 Java 项目中^[3],文中借鉴该方法将 Maven 引入仿真系统进行横向和纵向拆分,提高代码的鲁棒性及团队开发下的易管理性;邱祝文从系统应用角度出发,提出基于 Redis 的分布式缓存系统架构^[4],在该研究的基础上合理使用 Redis 集群,完成摄像头拍摄模拟模块的设计,并结合 Flume 框架^[5-7]保证分布式日志稳定可靠收集的需求;薛亚鹏提出“基于实时日志系统的海量日志服务平台设计与实现”^[8],魏彬提出“基于分布式日志系统的数据云服务平台设计与实现”^[9],均为日志收集提供思路,具体实现过程选用稳定可靠的 Hadoop^[10-14]框架完成日志收集。

1 仿真模拟系统模块划分

根据高并发模拟^[15-16]下计算机性能充分利用原则,以功能对计算机资源消耗程度为度量对项目进行横向和纵向的拆分,将系统划为 4 个模块:交通元素模拟模块、车辆模拟运行模块、摄像头拍摄模拟模块和日志收集模块,通过不同模块在不同服务器中的部署运行完成各自的功能。以下为系统各功能模块分析说明。

1.1 交通元素模拟模块

城市交通元素模拟模块是该系统的基础模块,将从需求分析、输入和输出进行阐述。

(1)需求分析。

该模块对城市交通运行中的“城市”、“交通卡口”、“摄像头”、“机动车”四个交通元素的模拟进行需求分析。表 1 是对该模块的需求划分。

表 1 城市交通元素模拟模块

元素名称	描述
城市的模拟	对呈正方形分布的城市从城市的宽(CITY_WIDTH)和高(CITY_HEIGHT)两要素进行合理的抽象和模拟
交通卡口的模拟	对虚拟城市中的所有街道(呈现井字分布)进行合理的抽象和模拟
摄像头的模拟	①对城市中所有的 10 000 台摄像头对象进行抽象和模拟 ②对 10 000 台设备分散架设在 2 000 个卡口对象的架设关系进行抽象和模拟 ③将所有摄像头按照所架设位置信息分组,以 JSON 数组的形式存储到内存数据库 Redis 中
机动车的模拟	①对城市中所有的 200 000 辆机动车对象进行抽象和模拟 ②车辆在城市道路中行驶过程进行合理的抽象和模拟

(2)输入和输出。

系统出于性能考虑对项目进行横向和纵向的拆分,拆分后该模块主要负责对系统中实体类的抽象,没有具体的数据输入和输出。

1.2 车辆模拟运行模块

车辆模拟运行模块是系统的核心模块之一,将从需求分析、输入和输出两个角度展开阐述。

(1)需求分析。

车辆模拟运行即对现实生活中城市道路上 20 万辆机动车同时运行过程的模拟。系统中,该模块需要对现实城市中道路中的车辆对象进行模拟,合理地将 20 万辆车分布到所有街道中,并对车辆在道路中的实时随机行驶及进过卡口这一事件进行模拟,同时保证 20 万级别的超高并发度,实现对现实生活中 20 万辆车在城市道路中并发行驶的模拟。表 2 为车辆模拟运行模块的需求划分。

表 2 车辆模拟运行模块需求划分

需求名称	描述
车辆模拟	①对现实城市中道路中的车辆对象进行模拟 ②合理地将 20 万辆车分布到所有街道中,完成对现实情况下城市中车辆随机分布行驶特点的模拟
车辆行驶行为模拟	①对单位时间内车辆的行驶过程进行模拟 ②对车辆每行驶一步返回的位置信息进行是否经过卡口做判断,经过卡口则需要生成行驶日志并向下层模块发送该日志
超高并发度行驶模拟	对 20 万辆车在城市中同时运行进行模拟

(2)输入和输出

该模块以 20 万辆机动车基本信息、2 000 个卡口基本信息为模块输入值,以携带有车辆悬挂牌照、车辆运行方向、经过卡口的位置等信息的海量级车辆行驶日志为输出值。

1.3 摄像头拍摄模拟模块

摄像头拍摄模拟模块是系统的核心模块之一,将从需求分析、输入和输出两个角度展开阐述。

(1)需求分析。

摄像头拍摄模拟即对现实生活中监控摄像头拍摄经过车辆的信息过程的模拟。在系统中,该模块负责对上层模块发来的短时间海量的车辆行驶日志进行处理,从车辆行驶日志中获取车辆所经过的卡口位置信息,最终获取所拍摄的摄像头信息并追加到车辆行驶日志中,成为摄像头拍摄日志,完成对现实生活中摄像头抓拍经过卡口的车辆牌照、抓拍时间、车辆行驶速度等拍摄过程的模拟。表 3 为摄像头拍摄模拟模块的需求划分。

表3 摄像头拍摄模拟模块需求划分

需求名称	描 述
车辆行驶日志处理	对上层模块传来的车辆行驶日志的位置信息进行提取 ①通过位置信息从存有摄像头对象的 Redis 内存数据库中获取该位置上所有的摄像头对象 ②从该位置上所有摄像头数组中,随机选中一台作为拍摄设备,并将摄像头基本信息追加到车辆行驶日志中,成为车辆行驶信息拍摄日志
生成摄像头拍摄日志	
向下层模块发送日志	将生成的摄像头拍摄日志发送到日志收集模块

(2)输入和输出。

该模块以上层模块传来的携带有车辆悬挂牌照、车辆运行方向、经过卡口的位置等信息的海量级车辆行驶日志及 Redis 中相应的摄像头对象 JSON 数组为输入值,以携带有车辆悬挂牌照、车辆运行方向、拍摄地点、拍摄时间、卡口 id、卡口位置信息、摄像头 id、摄像头基本信息等要素的摄像头拍摄日志为输出值。

1.4 日志收集模块

日志收集模块是系统的核心模块之一,将从需求

分析、输入和输出两个角度展开阐述。

(1)需求分析。

该模块负责对上层模块(分布式集群)发来的海量级的日志数据进行高效、可靠的收集及在分布式文件系统中的合理可靠存储。表4为日志收集模块的需求划分。

表4 日志收集模拟需求划分

需求名称	描述
海量级日志 高效可靠收集	对上层海量级日志数据进行可靠、高效、 实时收集
海量级日志数据 合理可靠存储	将实时收集到的海量日志数据在分布式 文件系统中进行合理的存储

(2)输入和输出。

该模块以上层模块(分布式集群)传来的日志数据为输入值,以分布式文件系统中日志文件为输出值。

2 仿真模拟系统网络架构设计

结合系统需求分析及项目架构设计对系统整体进行网络架构设计。系统网络架构如图1所示。

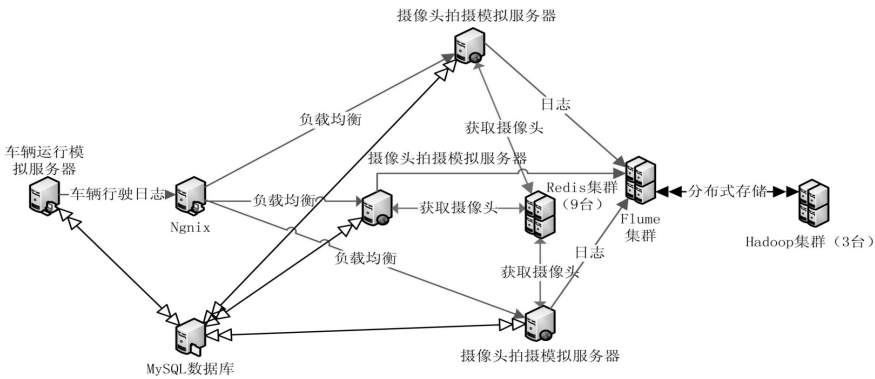


图1 车辆模拟运行及日志收集子系统网络架构

根据各个 Maven 子项目和各个模块的特点,将车辆模拟运行项目独立部署在一台服务器,使服务器性能得到最大发挥。由于摄像头拍摄模拟模块要抗击上层车辆模拟运行模块发来的海量级 Http 压力的特点,将摄像头拍摄模拟项目分别部署在三台服务器上,结合 Nginx 服务器的使用,实现负载均衡的效果,抗击上

层模块超高并发的 http 请求,保证摄像头拍摄模拟的性能。根据摄像头拍摄模拟模块高频率进行数据库 IO 操作的特点,使用 Redis(内存级数据库)并结合哨兵及分片机制来保障摄像头拍摄模拟服务器集群高频率的数据库 IO 操作。Redis 集群架构如图2所示。

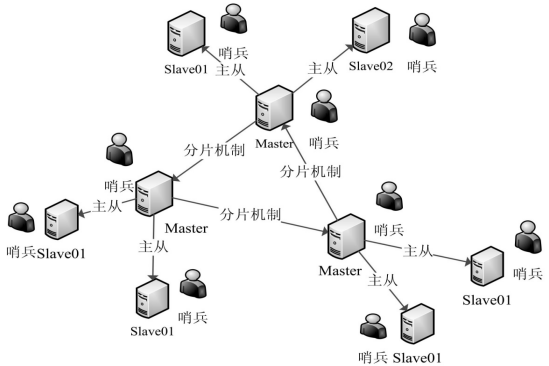


图2 Redis 集群架构

考虑到摄像头拍摄模拟模块分布式集群的特点,选用 Flume 并采取多级流动的结构保证摄像头拍摄模拟模块分布式日志稳定可靠收集的需求。在存储方面,考虑到该系统分布式集群输出海量级别的日志的特点,选用稳定可靠的 Hadoop^[8-12] 集群(3台),并将日志文件存储在分布式文件系统上。Flume^[14-16] 及 Hadoop 架构设计如图3所示。

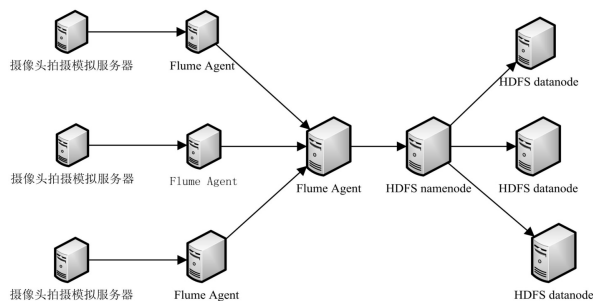


图3 Flume 及 Hadoop 架构设计

3 仿真模拟系统关键模块实现

针对该系统的关键功能模块给出具体的实现方式。

3.1 车辆模拟运行模块实现

车辆模拟运行功能即对20万辆车在道路中实时运行进行模拟。涉及的类包括 RunController 类、CarTask 类、HttpClientThread 类、HttpThreadFactory 类、SystemConfig 类、Car 类、HttpClientService 类、Position 类等。

(1) RunController 类。

负责车辆模拟运行的准备工作,包括 Runnable 任务创建、线程池创建等。由于车辆模拟运行功能的性能完全取决于线程并发性能及合理的调度,针对高并发的要求,对线程、线程池、对象复用做了专门的设计。由于篇幅原因,这里只对高并发需求下的线程相关设计的实现进行描述。

(2) CarTask 类。

实现自 Runnable 接口,为每一辆车创建一个 CarTask 对象,在 run() 方法中,执行对应的 Car 类对象的 move 方法,模拟车辆行驶一步。同时,根据 move() 方法返回的信息判断车辆当前步骤是否经过了卡口,如经过卡口,则使用当前线程中的 HttpClient 对象,向下层模块发送车辆行驶日志。日志格式如下:

①内容。

车辆 id | 真实车牌 | 悬挂车牌 | 当前拍摄时的坐标_X | 当前拍摄时的坐标_Y | 车辆上一步坐标_X | 车辆上一步坐标_Y | 拍摄时运行方向 | 车辆被拍摄的坐标_X | 车辆被拍摄的坐标_Y | 拍摄时间 (long) | 拍摄的摄像头 id

②编码。

```
id | registId | hangId | x | y | prex | prey | dir |
passedPortPosition_X | passedPortPosition_X | timestamp |
sensorId
```

HttpClientThread 类是为了解决20万级的数量下 HttpClient 对象的复用而设计的自定义线程类。将 HttpClient 对象从 CarTask 中抽出,封装在了每个线程中,大大提高了对象的复用及项目稳定性。

ScheduledExecutorService 属性 (RunController 类属性) 是为了针对车辆需要循环执行行驶步骤的要求而选用 Executors 框架中的定时调度线程池类对象。

addTasksToPoll() 方法是将200 000个 Runnable 任务 (CarTask) 提交至线程池,线程池会定时将任务队列中的队列提交给线程执行,保证每个任务执行的间隔相同。

至此,每个车辆以一定的时间间隔执行一步模拟行驶,如果经过卡口,则向下层模块发送车里行驶日志。车辆运行模块涉及所有功能均已实现。

3.2 摄像头拍摄模拟模块实现

摄像头拍摄模拟模块具体包括卡口模拟、摄像头模拟、摄像头拍摄模拟三个功能。以下将从这三个功能展开介绍。

(1) 卡口及摄像头模拟。

由于摄像头拍摄模拟模块采用分布式集群架构,摄像头拍摄模拟模块需要高效的根据上层模块传来行驶中车辆经过的卡口位置信息高效获取到对应的摄像头信息,因此选用内存级数据库 Redis,搭配哨兵及分片机制搭建本功能。同时将100 000台摄像头中位置相同的摄像头合并为一个集合,以位置信息为键,集合为值的形式存储在内存级数据库 Redis 集群中。

(2) 摄像头拍摄模拟。

摄像头拍摄模拟功能即摄像头拍摄模拟模块接收上层发送的车辆行驶日志,根据日志中的位置信息从 Redis 中获取相应的摄像头信息追加到日志中,成为拍摄日志,发送给下层模块。拍摄模拟功能时序如图4所示。

MessageHandlerController.handle(String carRunInfo) 方法负责接收到来自上层模块传来的车辆行驶日志,提取出车辆经过的卡口位置信息,内部调用 MessageHandlerServiceImpl 类对象的 getSensorIDByPocation(Position pos) 方法根据位置信息获取拍摄的摄像头 id。

MessageHandlerServiceImpl.getSensorIDByPocation(Position pos) 负责摄像头拍摄模拟的具体实现,内部通过 JedisCluster 实现类与 Redis 进行相应的通信。

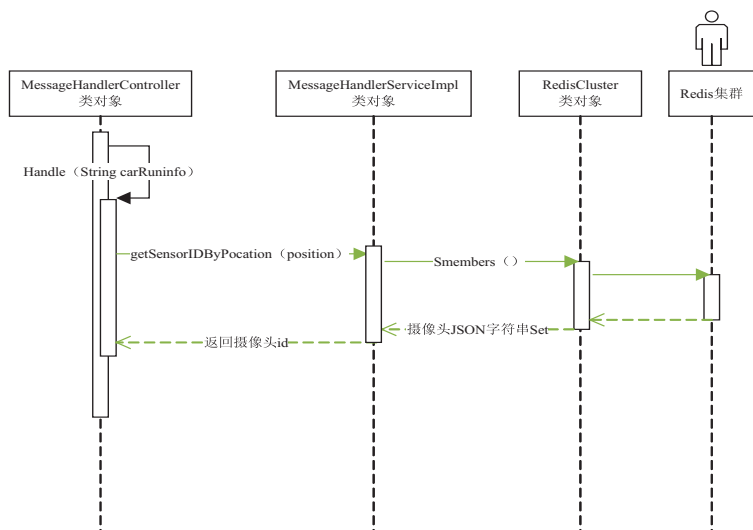


图4 拍摄模拟功能时序

jedisCluster.smembers()方法负责和Redis集群进行交互,借助本方法根据位置信息获取所有该位置上所有的摄像头的JSON格式字符串。从获取到的集合中随机选中一个摄像头对象JSON字符串,并提取摄像头id返回。

3.3 日志收集模块实现

日志收集模块即对分布式的摄像头拍摄模拟集群产生的海量日志信息进行稳定可靠的收集,具体分为分布式日志收集和海量日志文件分布式存储两部分。分布式日志收集部分采用现阶段成熟的Flume分布式日志收集框架进行功能的搭建,采用多级流动的形式提高上层拍摄模块高并发情况下的收集性能。海量日志分布式存储部分采用成熟的Hadoop生态系统中HDFS分布式文件系统来实现。由于篇幅原因,对Hadoop集群搭建过程不做描述。

4 结束语

借助大数据技术,该系统的研究和实现,不仅实现了城市级别的车辆模拟运行和日志收集,可以为城市交通提供完整、相对贴合实际的运行场景,而且以此模拟系统为依据,可以方便车辆稽查,甚至调整道路交通运行方案。该系统为数据量较大的行业提供了一种解决问题的思路,用同样的方法可以为城市旅游、城市产业分布等进行模拟,进而调整旅游服务、产业设置等方案。

参考文献:

- [1] 王伟哲. 基于安全访问和集群技术的车辆稽查系统的实现[D]. 上海:复旦大学,2011.
- [2] RAHMAN A,DASH S. Big data analysis for teacher recommendation using data mining techniques[J]. International Journal of Control Theory & Applications,2017,10(18):95-

105.

- [3] 江日念,林霞,乔德新. Maven在Java项目中的引入及应用[J]. 电脑知识与技术,2013,9(21):4842-4847.
- [4] 邱祝文. 基于redis的分布式缓存系统架构研究[J]. 网络安全技术与应用,2014(10):52.
- [5] 周波. 基于Apache Flume的MR数据采集实施方案[J]. 电信科学,2018,34(S1):216-221.
- [6] SHREEDHARAN H. Using flume flexible,scalable,and reliable data streaming[M]. New York:O'Reilly Media,2014.
- [7] STEVE H. Apache flume:distributed log collection for hadoop[M]. Piscataway:Packt Publishing,2015.
- [8] 薛亚鹏. 基于实时日志系统海量日志服务平台设计与实现[D]. 北京:中国科学院大学,2017.
- [9] 魏彬. 基于分布式日志系统的数据云服务平台设计与实现[D]. 杭州:浙江大学,2013.
- [10] 黄华林,庞欣婷. 基于Hadoop的数据资源管理平台设计[J]. 计算机应用与软件,2018,35(7):329-333.
- [11] KORAT V G,DESHMUKH A P,PAMU K S. Introduction to Hadoop distributed file system[J]. International Journal of Engineering Innovation & Research,2012,1(2):230-236.
- [12] HE Jie,LAI Min. Research on key technologies of distributed Hadoop data mining in cloud computing platform[J]. Machine Tool & Hydraulics,2018,46(24):144-149.
- [13] 朱颢东,冯嘉美,张志锋. 基于Hadoop的大数据处理平台研究[J]. 华中师范大学学报:自然科学版,2017,51(5):585-590.
- [14] 李辉,王建文,叶明雯. 基于Hadoop的海量气象水文数据并发处理模型[J]. 计算机应用,2018,38(S2):187-191.
- [15] 陈芳洲,韦韬,邹维. 一种提高模拟超大规模TCP并发连接性能的方法[J]. 计算机工程,2003,29(4):110-111.
- [16] 赵军锋,庄怀东,赵亮,等. 大规模数据并发环境下视频监控系统的研究与实现[J]. 都市轨道交通,2016,29(6):11-15.