



Hadoop大数据技术助力山东交警实现智慧交通

道路交通作为“衣食住行”中的“行”和我们的日常生活息息相关。我们对道路交通的需求无非是安全和快捷。但是在机动车越来越多的今天，交通事故屡见不鲜，交通堵塞更是家常便饭，我们的需求显得有些奢侈。要改变道路交通的现状，我们不仅需要政府的宏观举措——增加道路建设、加强交通法规教育、发展公共交通等，更需要交管部门落到细节的管理。我国的机动车保有量极大（一个省的机动车数量在千万级别），道路交通还具有不受统一调度、行车轨迹复杂等特点，管理难度很高。为了提高管理能力，各地的交管部门纷纷部署了统一的监控系统，通过电子眼、传感器、测速器等设备对交通情况进行全天候的监控。

道路交通监控可以分为两类——观察型监控和分析型监控。观察型监控只需要监控设备能够观察到特殊状况，比如闯红灯、违停、越线、超速等违章行为，并将该状况发生的时间、地点和电子眼拍下的照片一同传送给监控中心，由交警进行状况的记录和处罚的通知。我国交管部门的观察型监控的使用已经相当成熟，对违章行为的捕捉率非常高，有效地降低了违章率。分析型监控就要复杂许多，常见的任务有套牌车分析、伴随车分析、碰撞车分析、黑名单车辆预警、旅行时间计算、道路流量统计等等。这些任务需要交通卡口不间断地记录所有经过车辆的过车数据，并且对这些数据进行查找、关联、比对等处理。因为记录条数多并且包含图像信息，过车数据的体量非常庞大，对监控系统的存储、查询和计算能力都提出了很高的要求。事实上，由于数据量过大，大多数交管部门采用离线分析进行分析型监控，也就是将一个周期内（比如一天内）全部的过车数据都存储起来后再对整个数据集进行计算。这种处理方式显然延时过高，监控系统在特殊状况发生很久以后才能将结果报告给交警。分析型监控的任务常常具有时效性，比如黑名单车辆通过某个卡口时，需要系统立刻捕捉到这一行为并通知卡口附近的交警前往拦截；再比如道路流量统计的目的是通知交警在某地交通流量过大时前去疏导。



离线分析的高延时使得交警无法对这些状况进行及时响应。

分析型监控的技术难点在于监控系统需要在不断产生的、海量的、格式多样的过车数据中快速发现问题，最好可以在问题发生时就发出预警。交管部门传统使用的数据库能力有限，无法满足对实时性的需求。所幸，这个难题恰好是近年来热门的大数据技术的强项。大数据技术以其分布式的计算方式尤其擅长对海量数据的快速处理。大数据发展到现在已经有相对成熟的技术来处理以下三种问题：复杂的批量数据处理、基于历史数据的交互式查询和基于实时数据的流处理。其中的流处理，顾名思义，是在数据产生并流入系统时就进行处理并马上得出结果，非常适合分析型监控中过车数据不断产生的场景和对实时性的需求。日前，山东交警就在公安部无锡所（交通管理科学研究所）的帮助下部署了基于大数据平台的缉查布控系统，利用流处理技术做到了实时的分析型监控。

山东省的机动车保有量在全国一直数一数二，2014年7月在全国率先实现了缉查布控系统的省级联网。省级的缉查布控系统要处理的不仅是省内各地市的监控信息，还有各地市之间交通的数据，因而数据量非常庞大，处理逻辑也远高于普通市级系统。山东交警在全省部署的上千个卡口每时每刻都捕获着大量包含图像和视频的过车数据，在高峰期可达每秒近万条，每月的数据量可达TB级别。山东交警缉查布控系统原先使用的Oracle数据库在处理如此大量的数据时捉襟见肘。部署了大数据平台后，数据量过大带来的存储、查询和计算方面的一系列问题都得到了解决。尤其是新系统中的流处理引擎成功地做到了对套牌车分析、伴随车分析、黑名单车预警、区间测速、旅行时间计算、道路流量统计等分析型监控的实时分析和实时预警。图1描述了山东交警新缉查布控系统的软件逻辑架构：

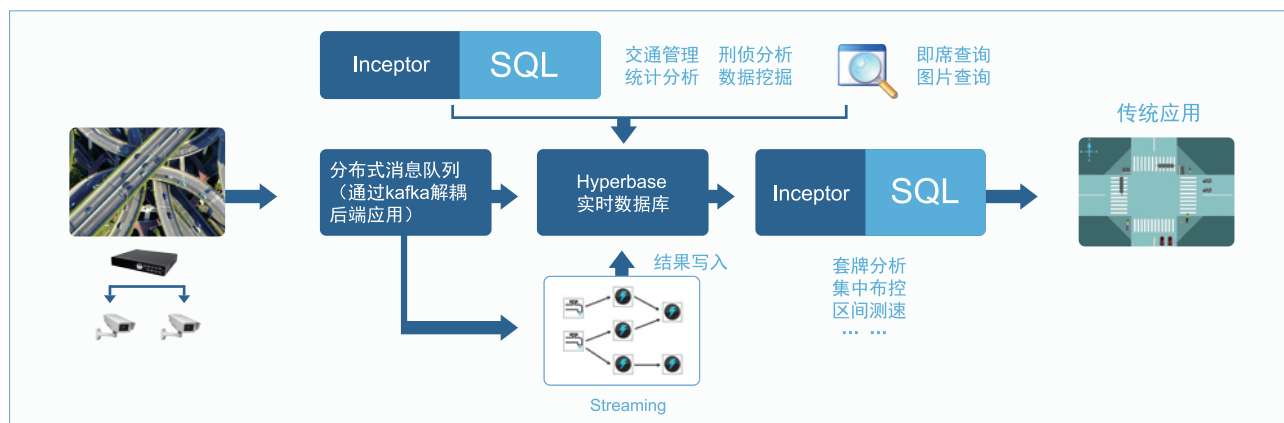


图1：软件逻辑架构

套牌车指拥有通过仿制、拼接等技术伪造与别的车辆相同的号牌的车辆。套牌车常被用作违法犯罪的工具，使案件难以侦察，给社会安全带来了很大的威胁。山东交警缉查布控系统对套牌车的分析根据“同一车辆在较短时间内不可能同时经过两个相距较远的卡口”的原理，将单个时间窗口内在全省产生的过车数据进行比对，在发现套牌车时立刻产生预警信息。信息一经交警确认，交管部门可以立刻通知卡口附近的交警前往拦截。大大提高了追踪套牌车的效率，可以帮助公安部门快速侦破套牌车所带来的违法犯罪行为。

伴随车指长时间结伴而行的车辆。当过车数据显示多辆车在多个卡口短时间内一起通过，缉查布控系统即会判定这些车辆为伴随车。犯罪团伙出动时，往往会驾驶多辆车协同作案。当交管部门确认某些路段会有犯罪团伙经过，对这些路段的伴随车分析可以迅速筛选出犯罪团伙所用车辆并通知公安干警前往拦截。

旅行时间计算和道路流量统计则可以帮助处理和预防交通堵塞。对车辆通过卡口的时间点进行计算，可以得到车辆在卡口之间所花的旅行时间。当系统发现大量车辆的旅行时间突然变长，则说明附近路段发生堵塞，缉查布控系统会立刻



发出预警，通知交警前往疏导。对各卡口过车数量的计数则可以计算出全省各条道路的实时流量，当道路流量大时，说明路上车辆很多，极易发生碰撞进而导致大面积交通瘫痪。所以缉查布控系统会在流量过大时通知交警前往待命，以便在状况发生后第一时间进行处理。在道路单向流量过大时，缉查布控系统还可以使用信号灯将流量小方向的部分车道临时调配给流量大的方向，实现流量均衡。

区间测速则可以根据一辆车通过两个或以上卡口的时间来计算出车辆在卡口间的速度，捕捉到超速行为。通过将实时生成的过车数据与车管所的黑名单比对，缉查布控系统可以发现包括未年检车辆、车主驾照吊销车辆、报废车辆等等并通知交警前往拦截。

使用了大数据流处理技术的缉查布控系统帮助山东交警建立了高度灵敏、快速响应的机制，大大提高了交管部门的管理能力。监控型分析之外，大数据平台还解决了山东交警海量数据的存储和查询问题。新的缉查布控系统中，机动车违法图像信息的保存周期从3个月延长到了24个月。交警可以根据车辆的颜色、车型、号牌等信息实时查询其历史行为、行车路线和车辆营运公司、驾驶人等关联信息，只花秒级时间便可从24亿条过车数据中的机动车号牌查询出精确结果和行车轨迹。值得一提的是新系统可以快速进行碰撞车分析。碰撞车并不是指真正在道路上碰撞的车辆，而是公安部门的一个术语，专指在两组或以上的过车数据中重复出现的车辆。我们可以用一个例子来解释碰撞车分析的意义。如图2所示：A地在某日凌晨2点、B地在同一日凌晨4点各发生一起盗窃案件，作案手法相同，公安机关认定为同一团伙所为。对A地附近卡口在该日凌晨2点左右和B地附近卡口在凌晨4点左右产生的过车记录进行比对，重复出现的车辆就有 犯罪嫌疑，公安部门可以迅速缩小侦察范围。



通过山东交警对大数据技术的利用，我们看到选择高性能的大数据处理平台可以大大提高交管部门的管理能力和效率。公安部无锡所帮助山东交警部署的新缉查系统底层采用了星环科技的分布式大数据处理平台Transwarp Data Hub（TDH）。TDH基于Hadoop，对海量数据的存储、查询和分析都很擅长，尤其是平台下的流处理引擎Streaming利用目前大数据处理领域中最快的计算框架Spark，拥有高吞吐、低延时的优点，使山东交警的缉查布控系统发挥了强大的作用。在交通监控系统中使用大数据技术不仅可以在突发事件中提高交警的响应速度，还可以为政府对道路建设和相关法规进行改进提供依据。例如，道路流量分析可以为公路局指出最需要新增和拓宽道路的地点，交通事故统计可以让交管部门在事故高发路段安装更清晰的警示标志或者重新设置限速等等。

交通拥堵和安全问题越发严重的今天，全国各省都在计划部署省级的缉查布控系统。河北、山西、重庆、湖南、浙江、甘肃等地区已完成相关项目中标，四川、安徽、辽宁、吉林、山西、江西、山东、湖北、青海等省市地区也已开始招投标工作。在省级缉查布控实现的同时，缉查布控的全国联网是大势所趋，届时，系统所面临的数据处理任务将更加艰巨。基于Hadoop的大数据平台扩展性极强，存储和计算能力都可以无限提升，在全国缉查布控系统中会发挥更大的威力。充分运用大数据技术，使道路建设、法规制定和事件处理配合无间，让交通管理变得更加“智慧”，道路交通时时处处安全、快捷的实现也就指日可待。

星环信息科技（上海）有限公司

📍 地址：上海市徐汇区桂平路481号18幢3层301室（漕河泾新兴技术开发区）

✉ 邮编：200233 ☎ 电话：4008 079 976

🌐 网址：www.transwarp.io

TRANSWARP
DATA HUB