



CHINA
DATA
ANALYST
SUMMIT

CDA 数据分析师
www.cda.cn

智慧交通大数据平台及应用

ADD YOUR POWERPOINT TITLE HERE

演讲人：曹正凤

跨界互联
数聚未来

第四届中国数据分析师行业峰会

CHINA DATA ANALYST SUMMIT

北京 中国大饭店 2017. 07

个人简介

曹正凤

统计学 博士

北京大数据协会 理事

经管之家（原人大经济论坛）大数据课程培训负责人
首发集团 智慧交通大数据中心筹备组负责人



博士论文：随机森林算法优化研究

出版专著：从零进阶！数据分析的统计基础

EI核心论文：Parameter Settings of Genetic Algorithm
Based on Multi-Factor Analysis of Variance

主持全国统计科学研究项目：基于大数据平台的我国宏观经济走势预测研究

从零进阶！数据分析的统计学基础



跨界互联
数聚未来

2017/7/17

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT



CDA 数据分析师
www.cda.cn

1

大数据团队概况及建设方案

2

智慧交通大数据平台

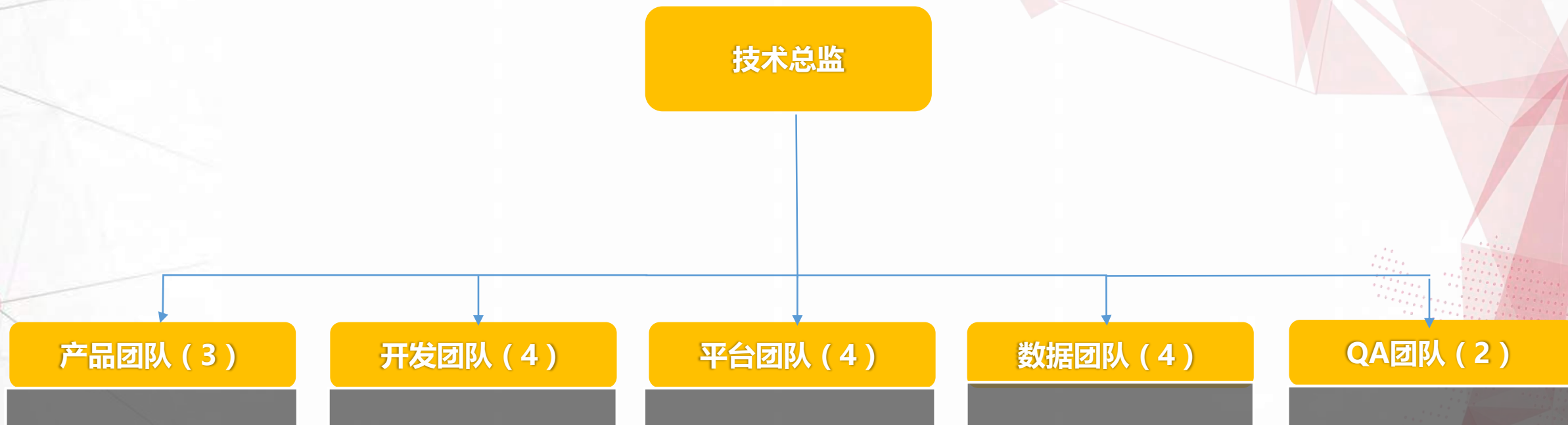
3

智慧交通大数据应用

2017

BigData Department

► 大数据部的组织架构



强大的技术架构

团队共**19**人

- ◆ 技术总监1人
- ◆ 专业数据挖掘分析师4人
- ◆ 大数据可视化工程师1人
- ◆ 需求分析师3人
- ◆ 总工程师1人
- ◆ 高级软件工程师5人
- ◆ 架构师1人
- ◆ 业务经理1人
- ◆ QA管理1人

2017

BigData Department

- ✓ 博士学历2人
- ✓ 研究生学历8人

智慧交通大数据中心建设方案（1+N战略）

大数据中心运维体系
运维管理制度、监控运维系统

智慧交通大数据应用类产品

交通大数据可视化（BI）

逃费稽查

OD分析与预测

视频结构化

入口自由流

...

智慧交通大数据平台

行业数据
规范与标准

行业数据
采集

行业数据
清洗

行业数据
规约

行业数据应用
支撑

数据分享
服务

行业规范数据存储

大数据基础平台

分布式数据存储

分布式计算引擎

平台运行监控

系统配置

管理门户

云计算基础平台（IaaS）

物理机、虚拟机、存储空间、网络等管理与租用服务

基础系统

虚拟化平台、操作系统、容灾备份

硬件：计算系统、存储系统、网络系统、安全设备

数据中心机房（租用和建设）

安全系统、安全管理、安全运维
云平台安全体系

一期建设的内容

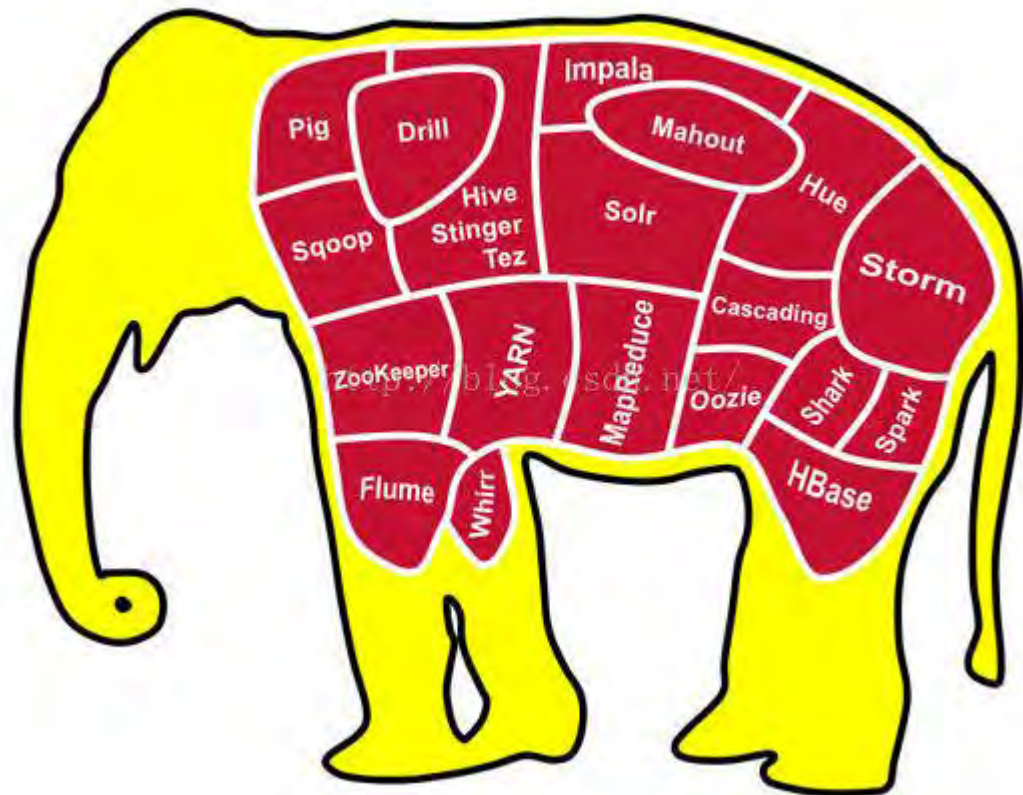


智慧交通大数据平台建设的目的和意义

人人都能进行大数据分析

大数据学习门槛高

Apache Hadoop Ecosystem



skill or knowledge

spark
mahout
JAVA
JDBC

effort or time

开源大数据平台的开发难度高

例如，黑屏命令行操作维护方式，增加了运维开发人员的工作难度、操作量，降低的工具的友好性。

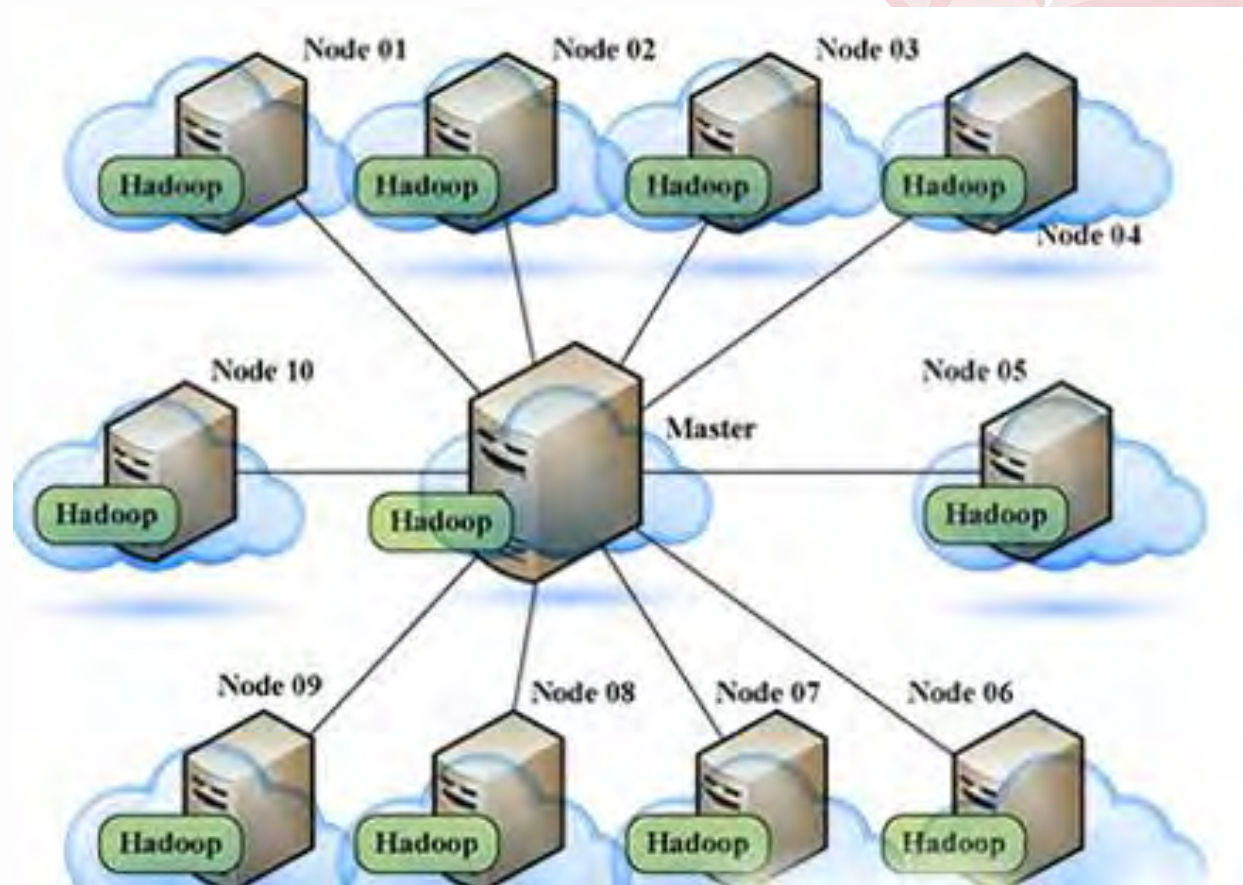
```
hduser@cloud01: ~/hadoop-2.2.0
4667 NodeManager
hduser@cloud01:~/hadoop-2.2.0$ ./bin/hadoop jar ./share/hadoop/mapreduce/hadoop-
mapreduce-examples-2.2.0.jar pi 5 10
Number of Maps = 5
Samples per Map = 10
Wrote input for Map #0
Wrote input for Map #1
Wrote input for Map #2
Wrote input for Map #3
Wrote input for Map #4
Starting Job
15/01/26 22:32:48 INFO client.RMProxy: Connecting to ResourceManager at cloud01/
192.168.75.101:8132
15/01/26 22:32:49 INFO input.FileInputFormat: Total input paths to process : 5
15/01/26 22:32:49 INFO mapreduce.JobSubmitter: number of splits:5
15/01/26 22:32:49 INFO Configuration.deprecation: user.name is deprecated. Inst
ead, use mapreduce.job.user.name
15/01/26 22:32:49 INFO Configuration.deprecation: mapred.jar is deprecated. Inst
ead, use mapreduce.job.jar
15/01/26 22:32:49 INFO Configuration.deprecation: mapred.map.tasks.speculative.e
xecution is deprecated. Instead, use mapreduce.map.speculative
15/01/26 22:32:49 INFO Configuration.deprecation: mapred.reduce.tasks is deprec
ated. Instead, use mapreduce.job.reduces
15/01/26 22:32:49 INFO Configuration.deprecation: mapred.output.value.class is d
```

开源大数据平台运维成本高

大数据集群数量众多的服务器不便于整体的部署与管理。

1. 资源监控
2. 组件部署
3. 资源分配
4. 辅助工具

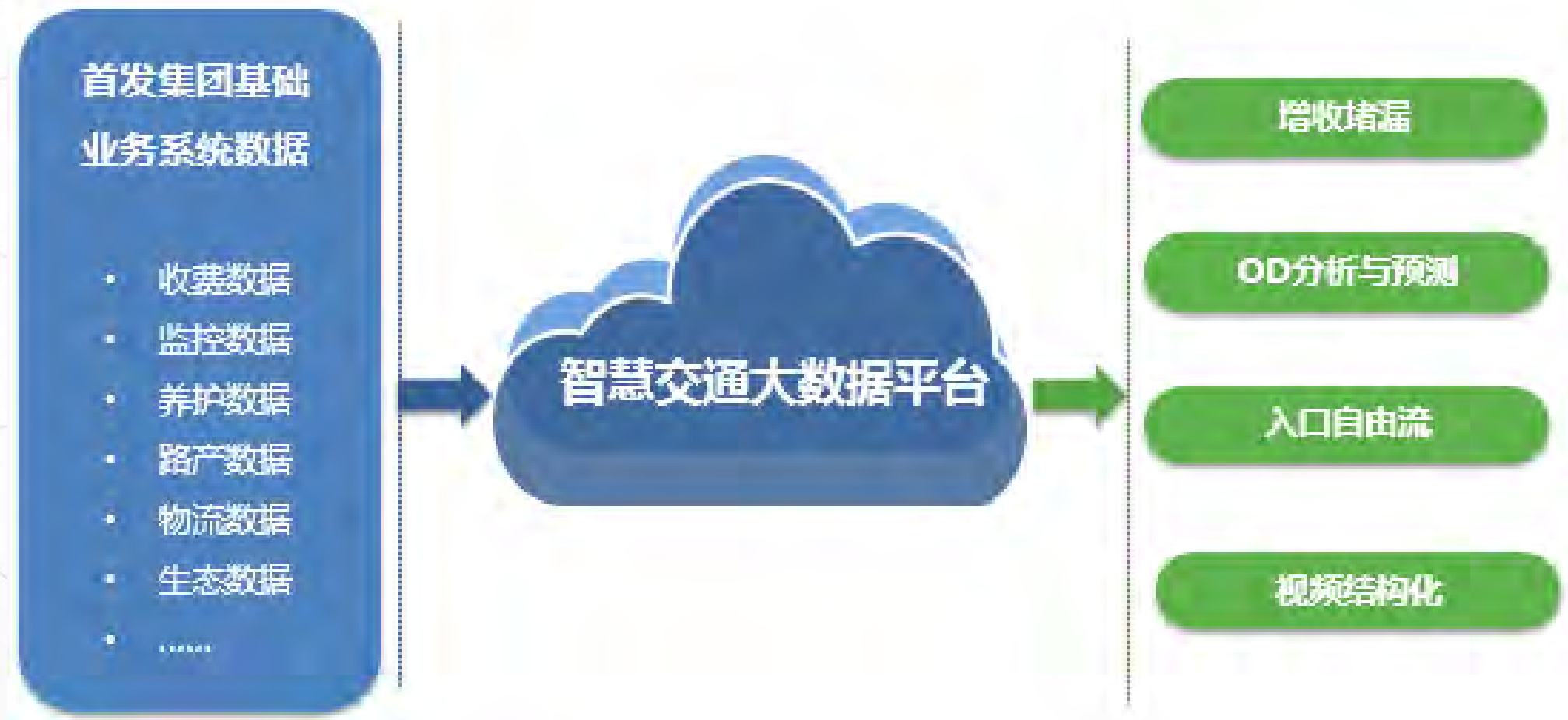
...



智慧交通大数据平台的功能特性

功能特性——各种异构数据的全方位汇集

我们



功能特性--拖拽式、图形化大数据分析环境，降低门槛和使用难度



• 功能特性——便于维护

自动部署

- 操作系统、LDP平台自动化部署功能
- 添加、批量移除，触发hadoop、hbase等服务的一系列处理

参数配置优化

- 图形化界面操作，配置过程简单、易操作；
- Hadoop基础组件（Hadoop-core、Hive、HBase）配置参数1000+；

服务器状态监控

- CPU、内存、网络、磁盘运行状态
- 超过阈值报警

任务监控

- 清晰的看到主机上运行了那些任务，任务的运行状态，是正常运行还是有故障
- 各个任务运行的内存、CPU的使用情况

• 功能特性——安全可靠

安全授权

- LDP平台可以控制数据访问，并对已通过验证的用户提供数据访问特权。

细粒度访问控制

- LDP平台支持细粒度的Hadoop数据和元数据访问控制；在服务器、数据库、表和视图范围提供了不同特权级别的访问控制，包括查找、插入等；允许管理员使用视图限制对行或列的访问，支持字段级别的访问控制；

基于角色的管理

- LDP平台通过基于角色的授权简化了管理，可以将访问同一数据集的不同特权级别授予多个组。

多租户管理

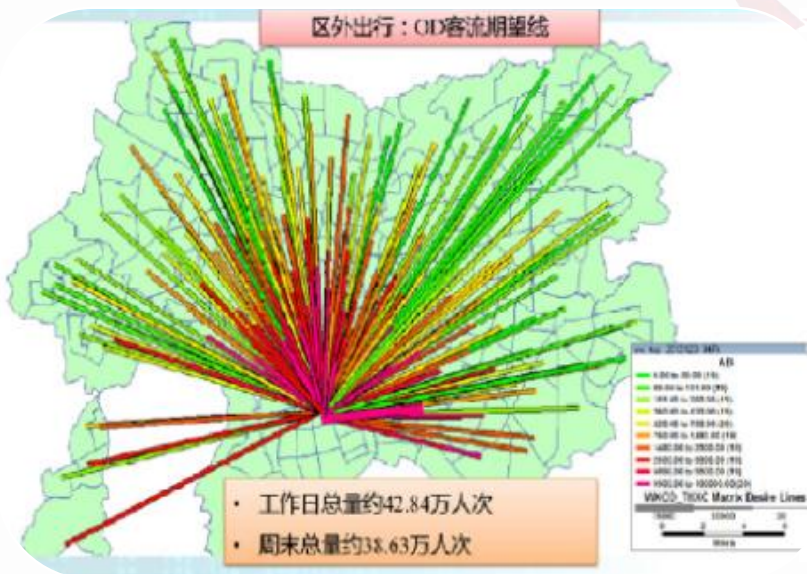
- LDP平台允许为委派给不同管理员的不同数据集设置权限。

机器级别的安全认证

- 对可信任的客户端提供认证，确保他们可以执行作业的相关操作。防止用户恶意冒充client提交任务。即使伪装成用户入侵到一个HDFS 或者MapReduce集群上，也无法读取HDFS上的数据。



综合稽查



BI展示



视频结构化

基于大数据平台的综合稽查

综合稽查

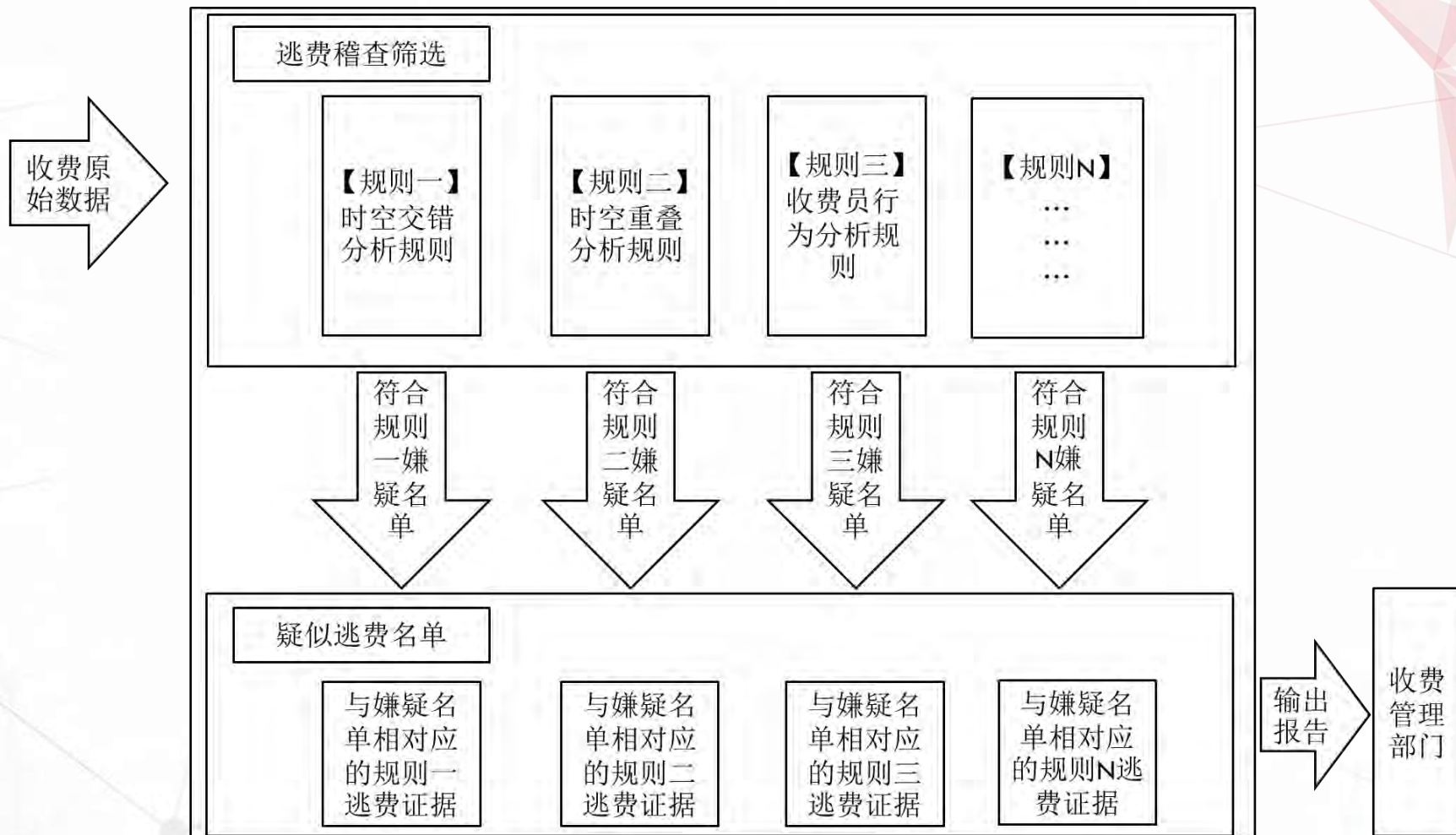
产品概况

针对当前高速公路的逃费问题，提出了一套基于大数据分析的高速公路防逃费稽查方案，本系统方案利用当前高速公路视频监控系统的可视化视频和图片信息结合通行信息，运用大数据存储和分析的手段，从海量历史数据中提取出疑似逃费车辆和黑名单的车牌、抓拍图片、通行视频和通行路径等证据信息，形成逃费证据链信息，该证据链信息可有效用于高速公路路政执法、公安执法，为打击高速公路偷逃费、嫌疑车辆定位提供了完善的稽查手段和证据信息。

主要功能

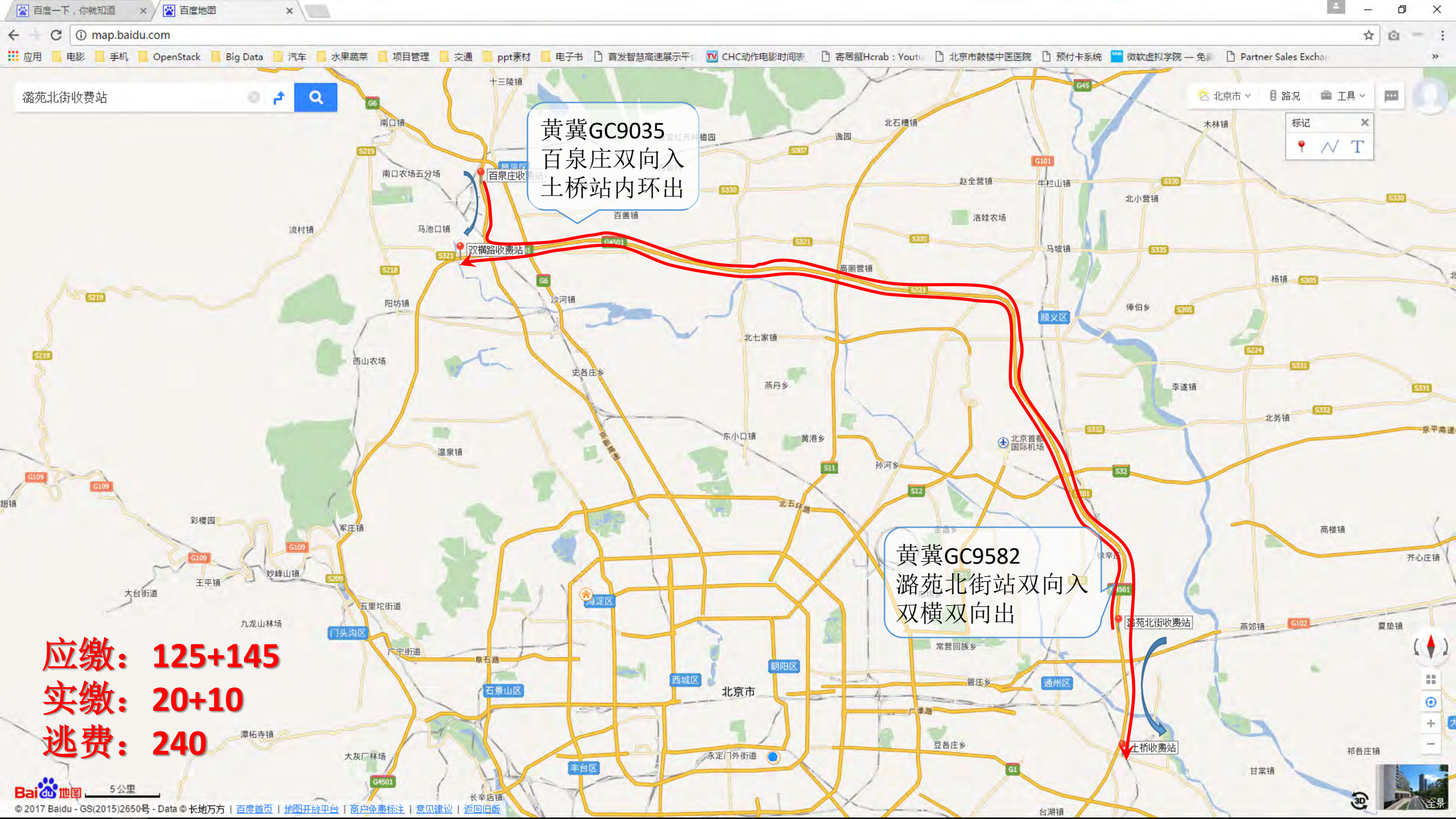
嫌疑车辆确认、黑名单管理、车辆特征管理、动态超时管理、数据稽查审核管理、证据链管理、车票及通行券路径查询。

综合稽查系统架构



异车倒卡处理流程



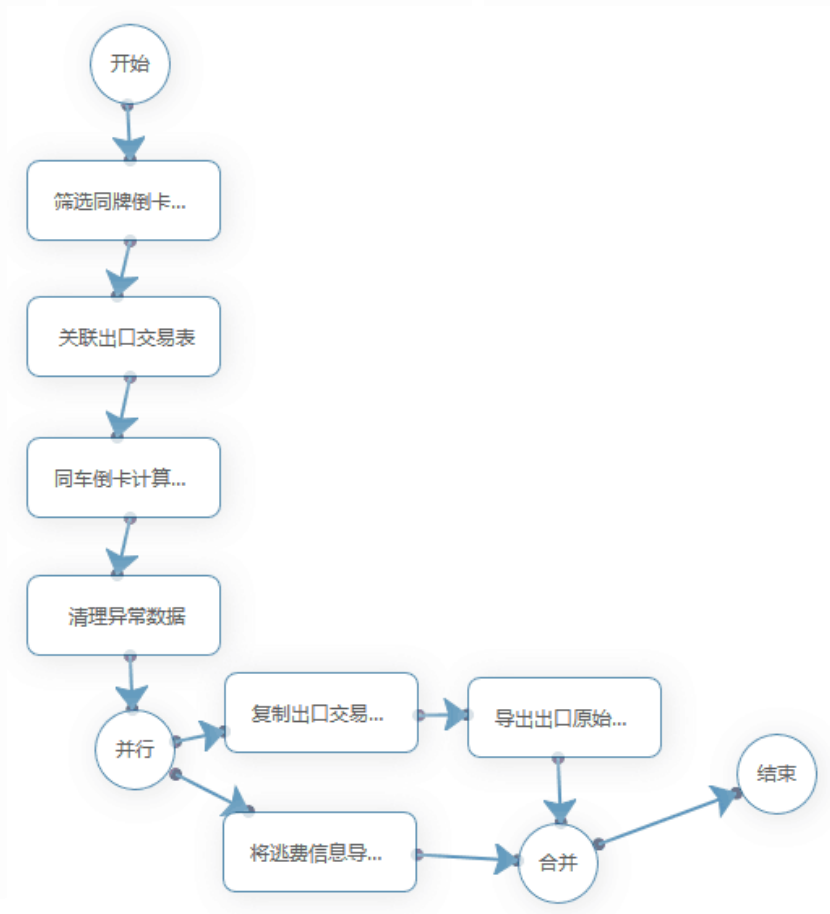


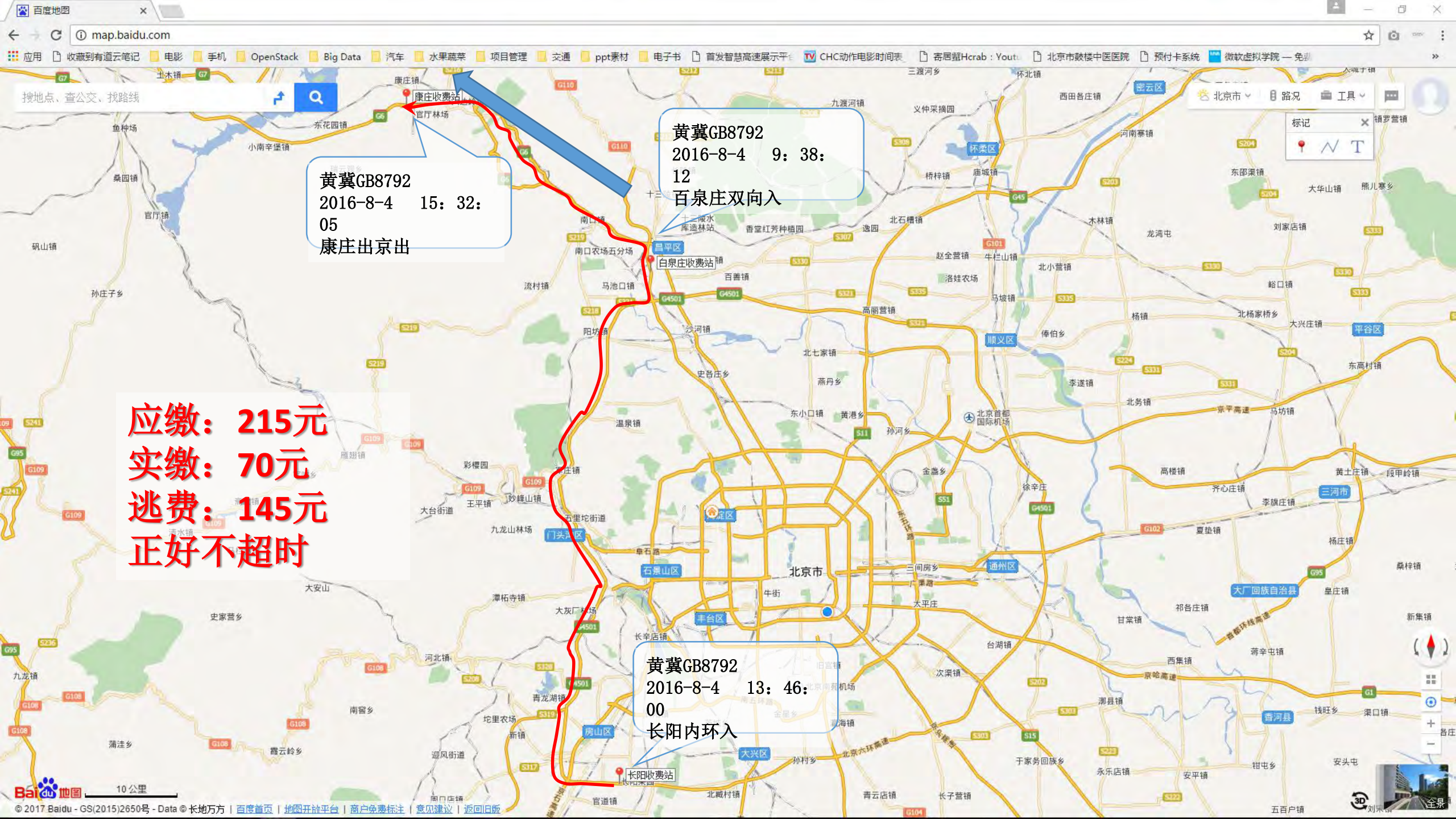
黄冀GC9035
百泉庄双向入
土桥站内环出

黄冀GC9582
潞苑北街站双向入
双横双向出

应缴: 125+145
实缴: 20+10
逃费: 240

同车倒卡处理流程





黄冀GB8792
2016-8-4 15: 32:
05
康庄出京出

黄冀GB8792
2016-8-4 9: 38:
12
百泉庄双向入

黄冀GB8792
2016-8-4 13: 46:
00
长阳内环入

应缴: 215元
实缴: 70元
逃费: 145元
正好不超时

应用之二：
视频数据
分析

车辆多维
特征提取

交通参数
提取

交通事件
检测

以图搜图

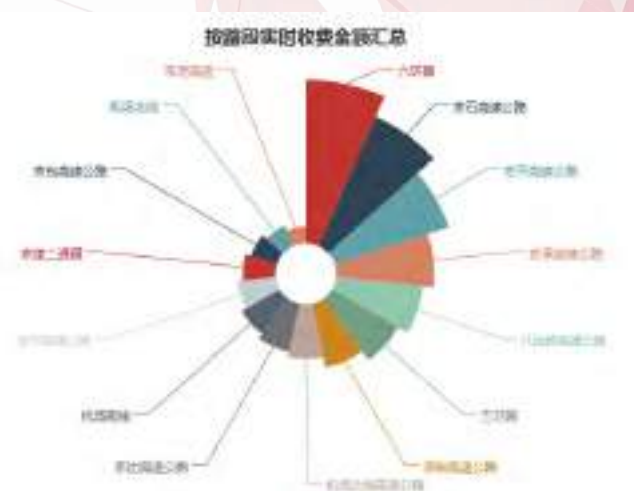
应用之三：基于大数据平台的实时BI展示



实时动态显示交通流量，时间间隔为2s



实时动态显示收费任务完成进度



实时动态显示按路段收费金额汇总

批量计算(按车型统计车流量)

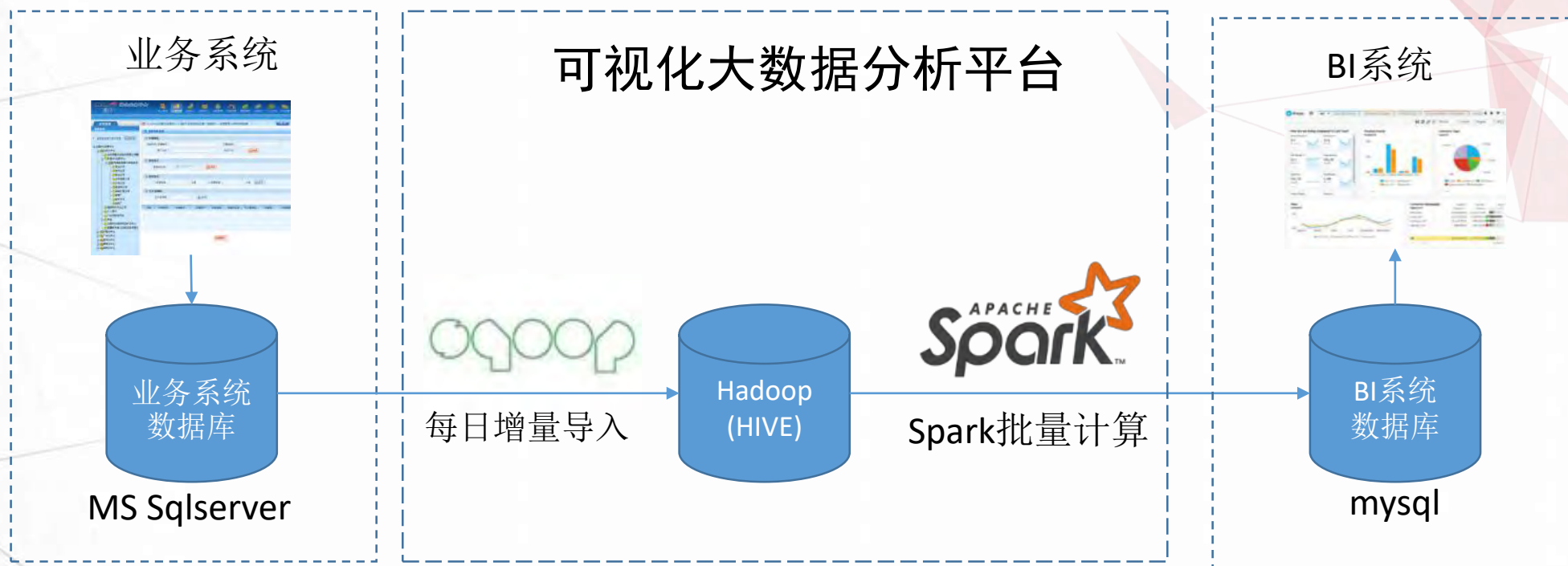
目的：实现每日从业务系统抽取交易数据，并按车型汇总交通流量。

批量计流程：

- 1、每日凌晨2点，Sqoop从分中心数据库(SqlServer)抽取数据到Hive
(由于车道数据上传到分中心有延迟，故2小时候启动导入)
- 2、启动spark任务进行统计计算，并将结果写入到Mysql
 - 按车型统计每小时交通流量和收费数据
 - 按车型统计每天交通流量

BI展示：JAVA 实现Web服务读取批量计算后的数据通过Web浏览器展现

架构设计



Sqoop每日定时导入数据
Impala结合SparkSQL实现统计运算

智慧高速

📊 监测板 📊 收费分析 📊 收费减免分析 📊 特殊事件分析

🔍 Search

实时收费监测

今日收入总金额：7352元



年累计收费总额
153556万元

本年累车流量
943489万辆

年累计减免总额
74329万元

实时交通量监控

今日车流总量：287辆



全路网收费月趋势

月收费总额：13803万元



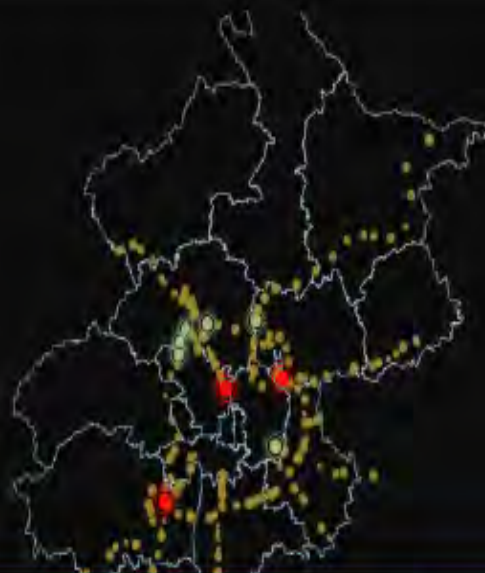
全路网减免收费趋势

月减免总额：1387万元



日期: 2016年8月12日
时间: 14时14分

● 畅通
● 正常
● 拥堵



拥堵指数:
收费增长指数:
流量增长指数:

收费率 (万元/分钟)



通行率 (万辆/分钟)



减免车通行比例 (%)



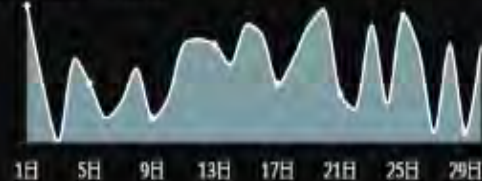
全路网交通量月趋势

月总流量：7608万辆



全路网减免交通量趋势

月总减免流量：452万辆





CDA 数据分析师
www.cda.cn

THANKS

跨界互联 数聚未来

第四届中国数据分析师行业峰会
CHINA DATA ANALYST SUMMIT