大数据交通运输行业统计应用实践及探索

Big Data Application and Exploration in Transportation Statistics

2019.10杭州　交通运输部科学研究院　陈建华

October, 2019 HANGZHOU

China Academy of Transportation Science　Chen Jianhua

中国拥有世界上最庞杂的交通运输系统

China has the most complex transportation system in the world

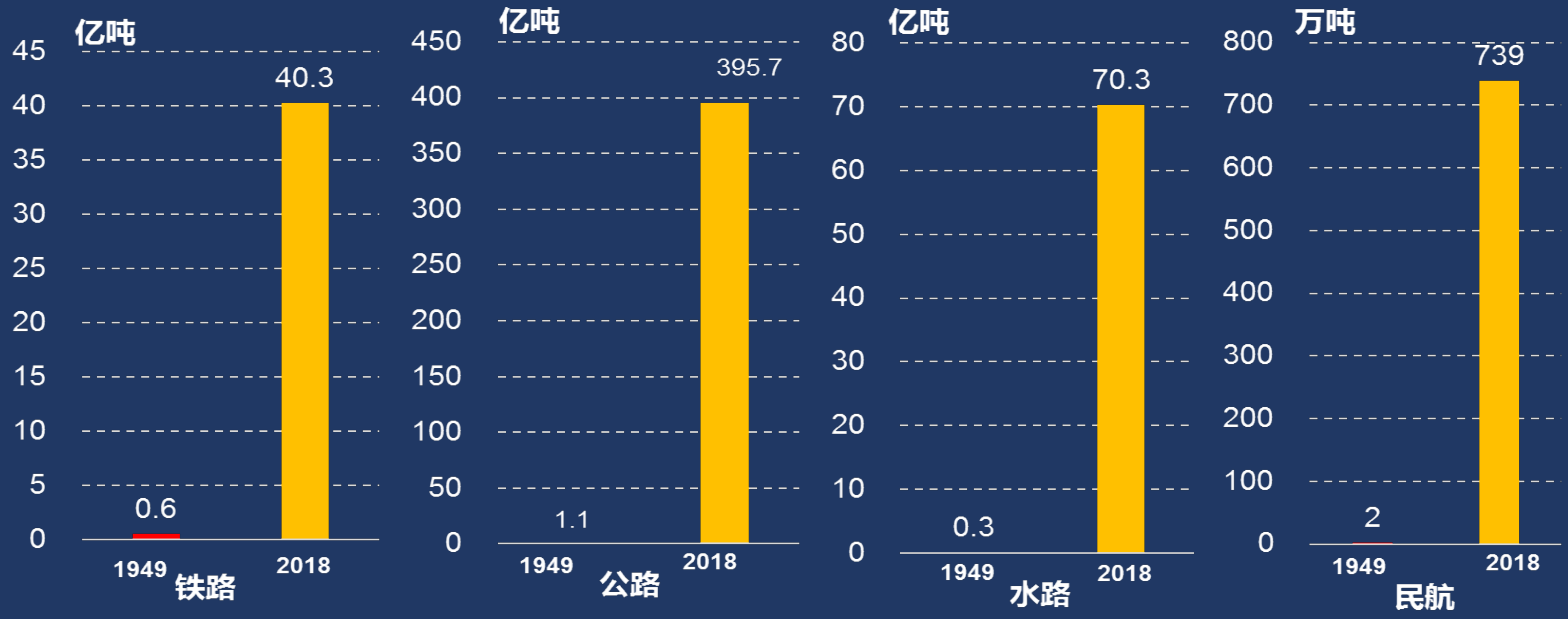
庞大的设施网络 Developed Infrastructure Network

- 铁路营业里程：13.1 万公里
- 公路线路里程：484.7 万公里
- 港口生产用码头泊位：2.4 万个
- Railway length— 131 thousand kms
- Highway length— 4847 thousand kms
- Port berth for transportation profits— 24000

活跃的市场主体 Active Market Entities

- 道路客运经营业户：3.4 万户
- 道路货运经营业户：569.9 万户
- 道路运输从业人员：2306.7万人
- Road passenger transport entity—34 thousand
- Road freight transport entity—5699 thousand
- Road transport employee—23067 thousand

旺盛的运输需求 Active Transportation Demand



- 营业性货运量1949-2018年间，年均增长8.5%
- During 1949-2018, the average annual growth rate of total cargo shipment is 8.5%

交通运输统计工作面临巨大挑战

Great Challenges for Transportation Statistics

数据需求时效性、频度越来越高

每日8:00前8个动态指标报李小鹏部长;

每旬旬后2日内关键指标报国办（高速公路货物运输量、全社会货物运输量）;

Needs for more timely and frequently data

8 key indicators should be presented to the minister of MOT, Li Xiaopeng before 8:00 am everyday

The freight shipment of expressway and the total transportation network should be submitted to General Office of the State Council in 2 days after every ten-day

交通运输统计工作面临巨大挑战

Great Challenges for Transportation Statistics

数据需求时效性、频度越来越高

每月10号主要指标报所有部领导

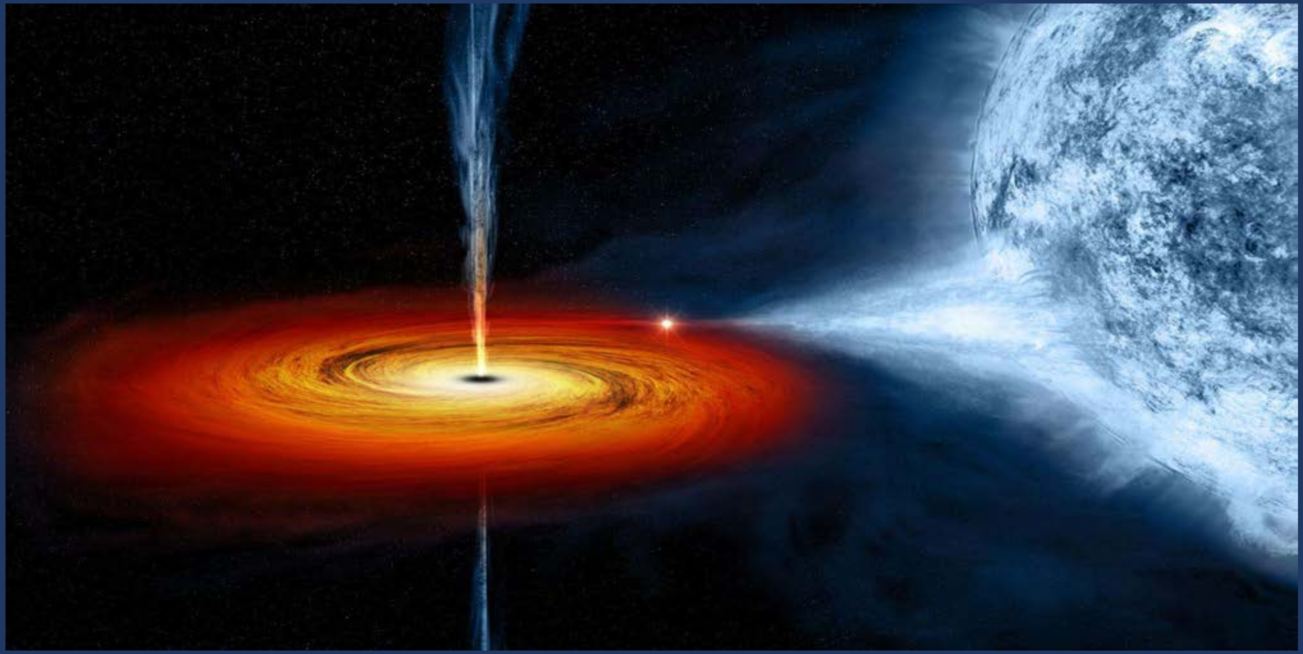
Needs for more timely and frequently data

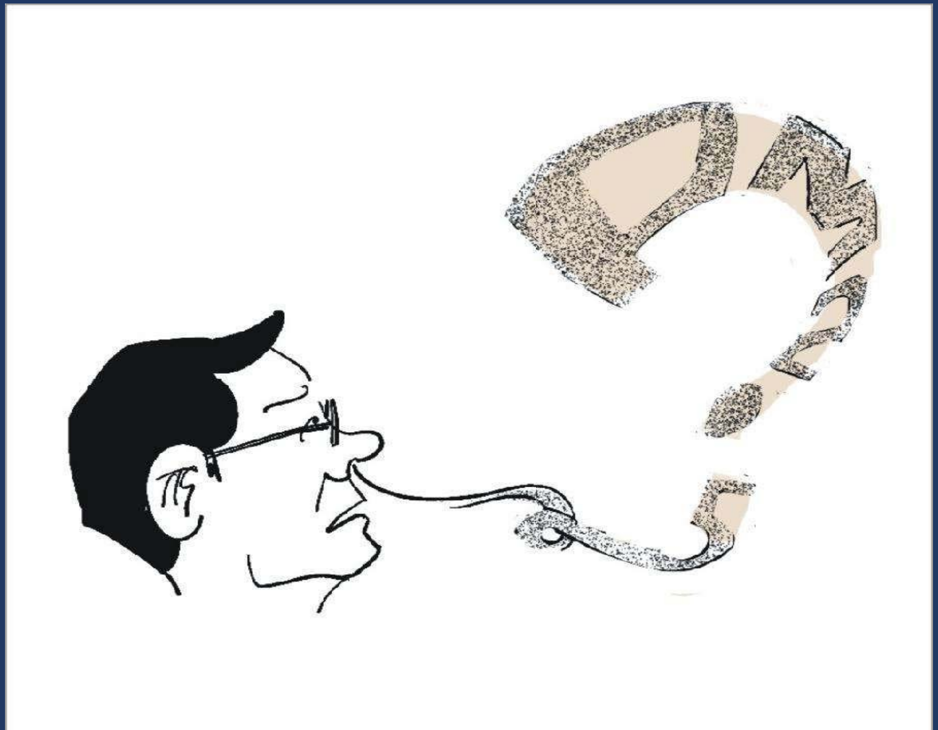
The composite indicators should be presented to all the leaders of MOT in ten days after the reporting month

交通运输统计工作面临巨大挑战
Great Challenges for Transportation Statistics

个别指标数据质量受到冲击和公众质疑

The particular data was influenced by subjective factors, thereby its quality was questioned by public





交通运输统计工作面临巨大挑战
Great Challenges for Transportation Statistics

市场主体基数大，集中度低

Large scale but low integration

客运经营：　个体经营业户2.2万户，占比64.7%

货运经营：　个体经营业户513.3万户，占比90%

Passenger transport entity：22 thousand individually owned entities accounting for 64.7% of the total

Freight transport entity: 5133 thousand individually owned entities accounting for 90% of the total

交通运输统计工作面临巨大挑战
Great Challenges for Transportation Statistics

基层统计工作可依赖的手段不足

Lack of effective measures for local statistics

简政放权、统计执法

Streamlining administration and delegating power

Law enforcement of statistics

主动寻求改革，不断追求创新
Calls for Reform and Innovation

顶层设计优化规范

Strengthen the top-level regulation

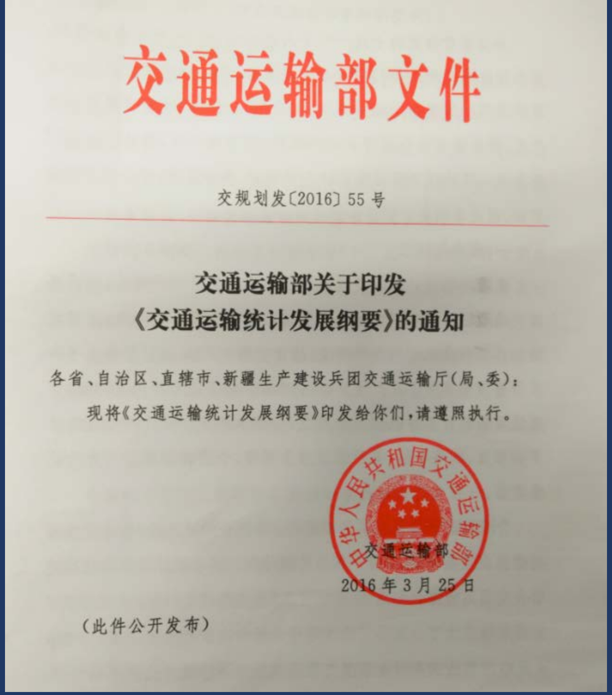

中华人民共和国交通运输部令

2018年第20号

《交通运输统计管理规定》已于2018年4月11日经第5次部务会议通过，现予公布，自2018年10月1日起施行。经国家统计局同意，1992年10月1日起实施的《公路、水路运输全行业统计工作规定》(交通部、国家统计局令1992年第36号)同时废止。

部 长

2018年7月23日


交通运输部文件

交规划发〔2016〕55号

交通运输部关于印发
《交通运输统计发展纲要》的通知

各省、自治区、直辖市、新疆生产建设兵团交通运输厅(局、委)：

现将《交通运输统计发展纲要》印发给你们，请遵照执行。

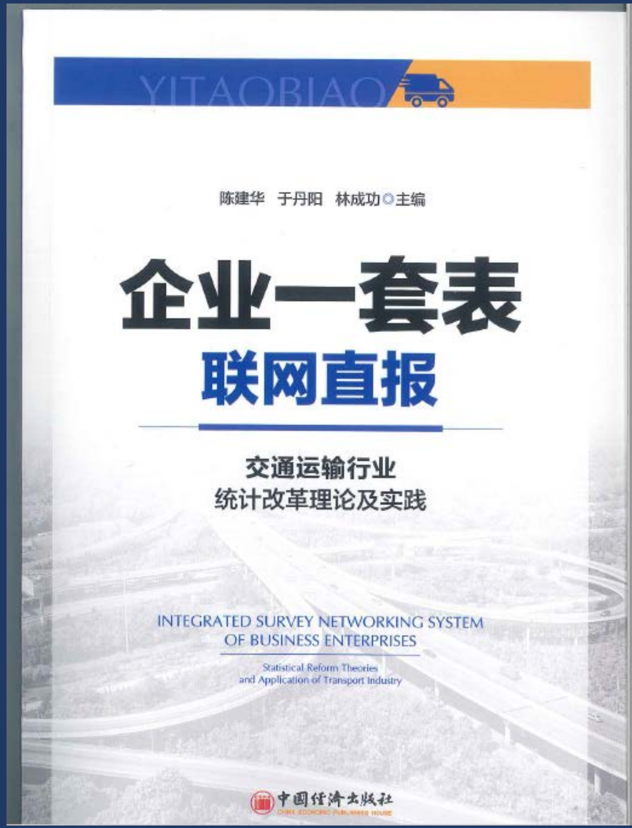
交通运输部

2016年3月25日

(此件公开发布)

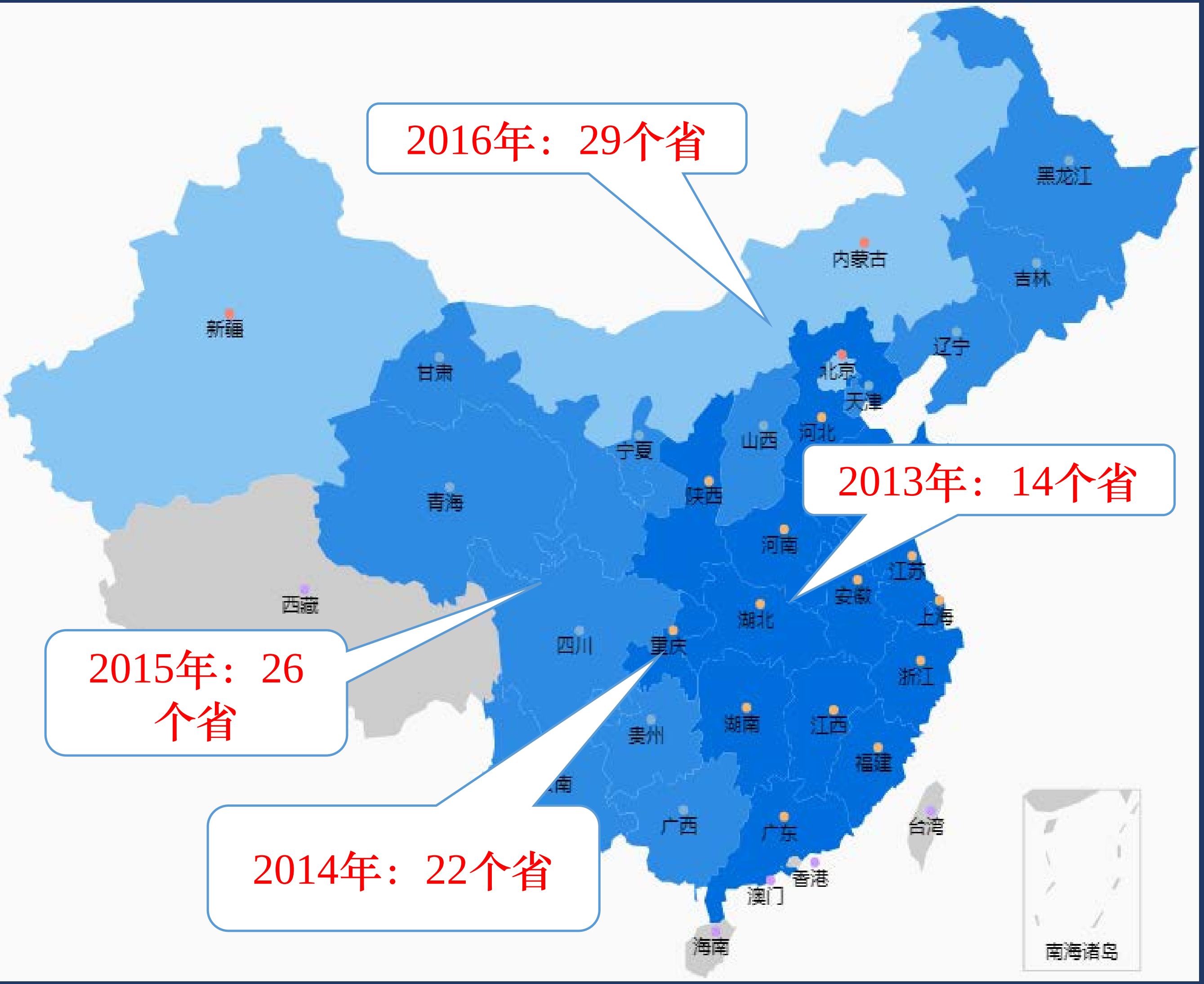
统计调查技术革新

Technical innovation of statistical surveys



高速公路运输量统计

Expressway Passenger and Freight Traffic



数据范围

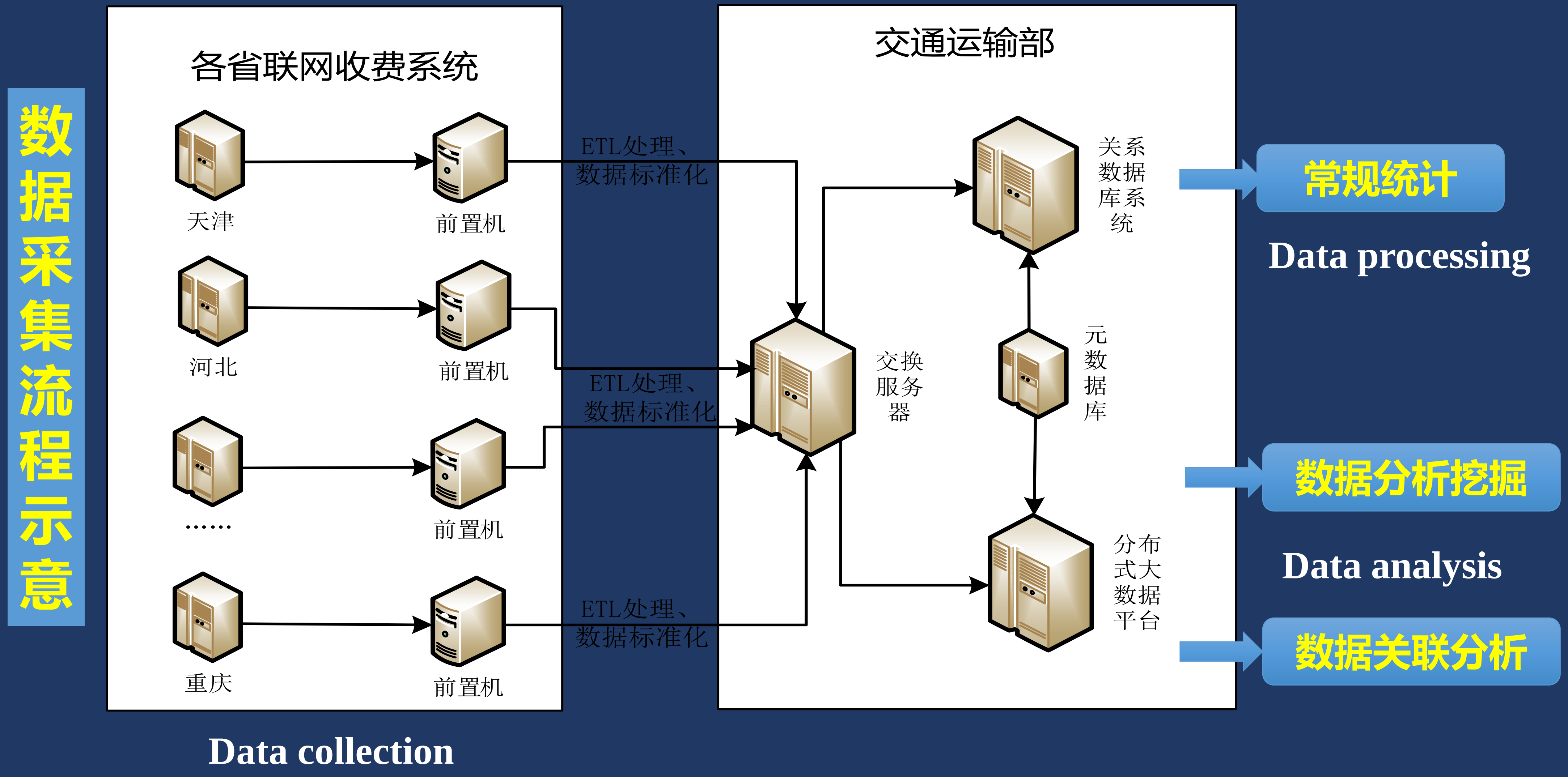
◆2013年：14个省份
◆2014年：22个省份
◆2015年：26个省份
◆2016年后：29个省份
（不含:西藏、海南）

2013, 14 provinces
2014, 22 provinces
2015, 26 provinces
Since 2016, 29 provinces
(Tibet, Hainan excluded)

Statistical Coverage

高速公路运输量统计

Expressway Passenger and Freight Traffic



高速公路运输量统计
Expressway Passenger and Freight Traffic

时效性
Time Requirement

1月
Every month

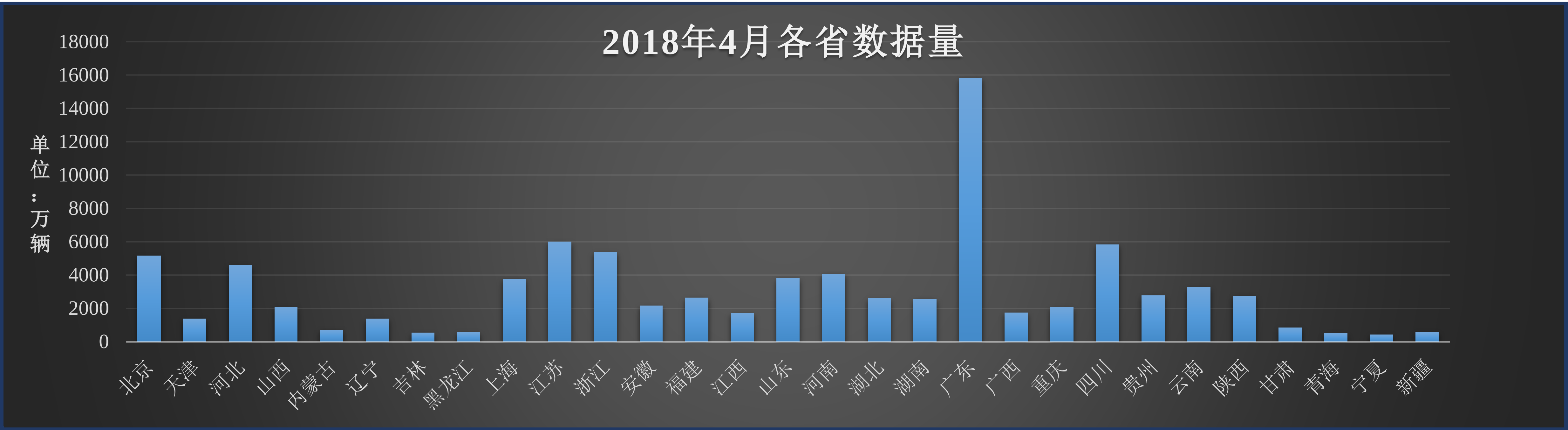
1旬
Every ten-day

15分钟
Every 15 minutes

高速公路运输量统计
Expressway Passenger and Freight Traffic

合计：收费站数量：9400个
高速公路里程：13.6万公里
数据容量：10TB
记录条数：400亿条

Toll station， 9400
Expressway length，136 thousand kms
Data capacity， 10TB
Total records，40 billion

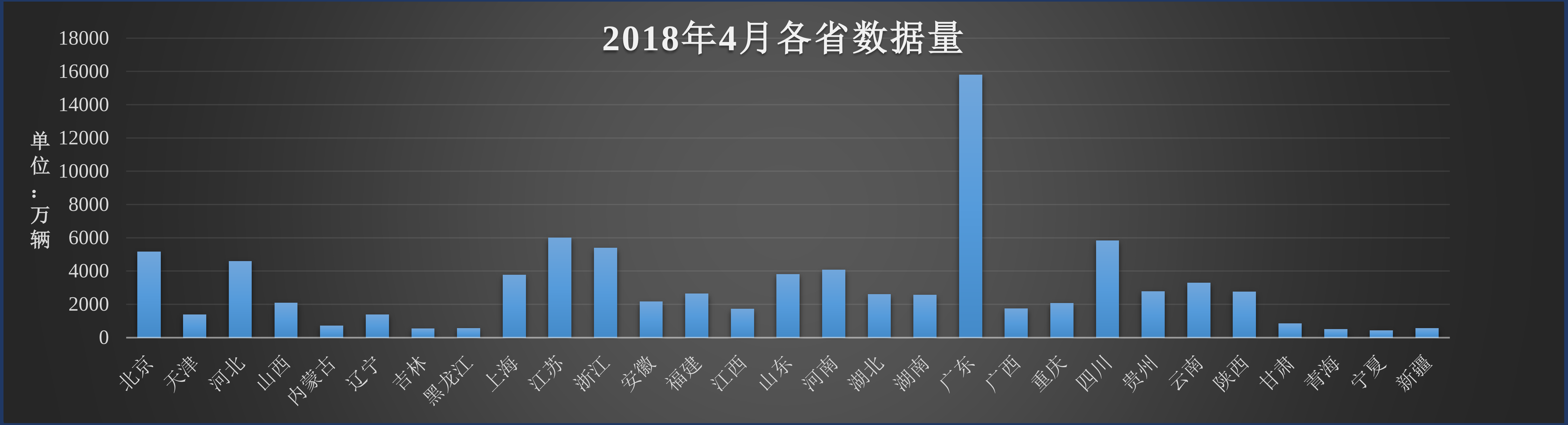


高速公路运输量统计
Expressway Passenger and Freight Traffic

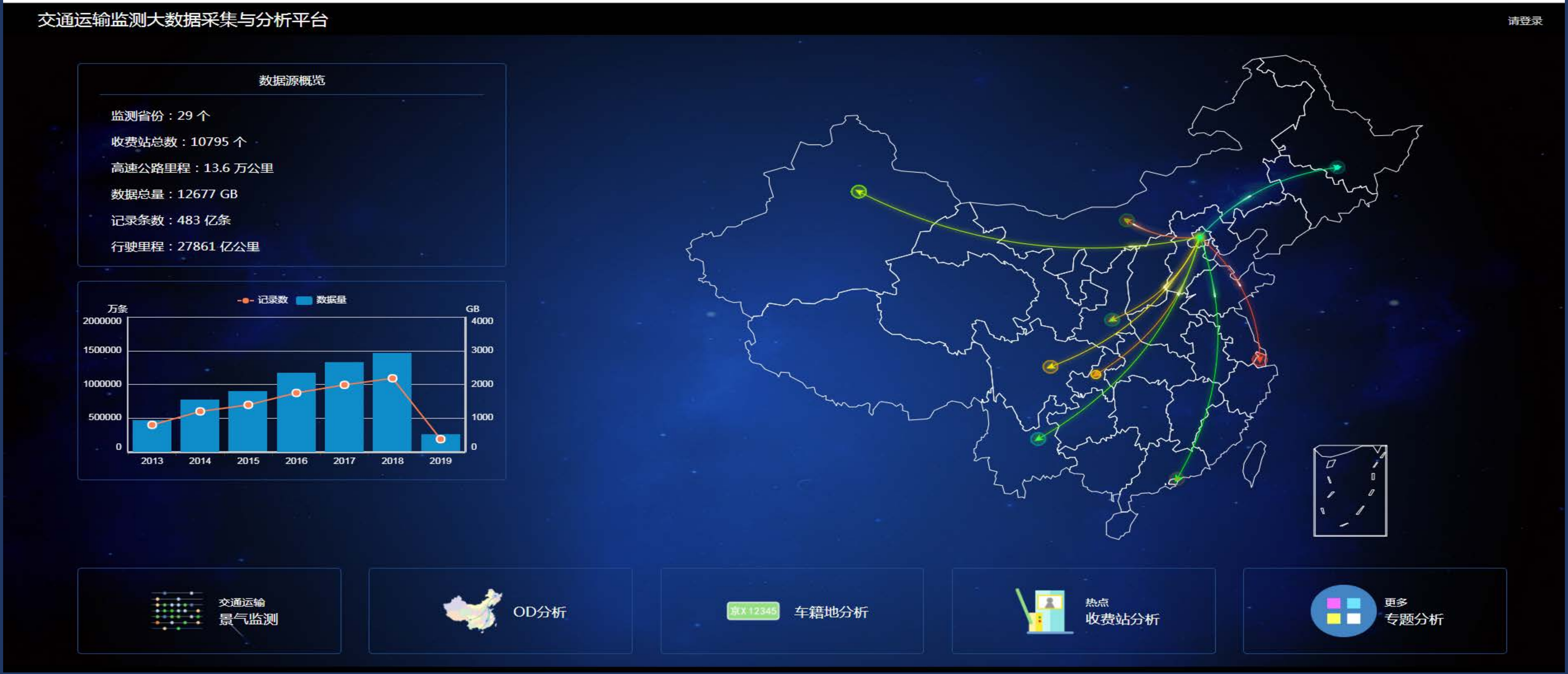
月均：数据容量：230 GB

记录条数：约9亿条

Average level per month
Data capacity ，230GB
Total records：about 0.9 billion



高速公路运输量统计
Expressway Passenger and Freight Traffic



高速公路运输量统计
Expressway Passenger and Freight Traffic

数据格式不一致
Inconsistent data format

数据内容不规范
Irregular data content

车型VEHICLE CODE：11、12、13、14、15
1 、2 、3 、4 、5

时间TIME：2018:03:01 07:40:10
2018.3.1 7.40.10

数据丢失
Data leakage

数据重复
Data duplication

部分日期或收费站数据漏报
Data leakage of specific date and stations

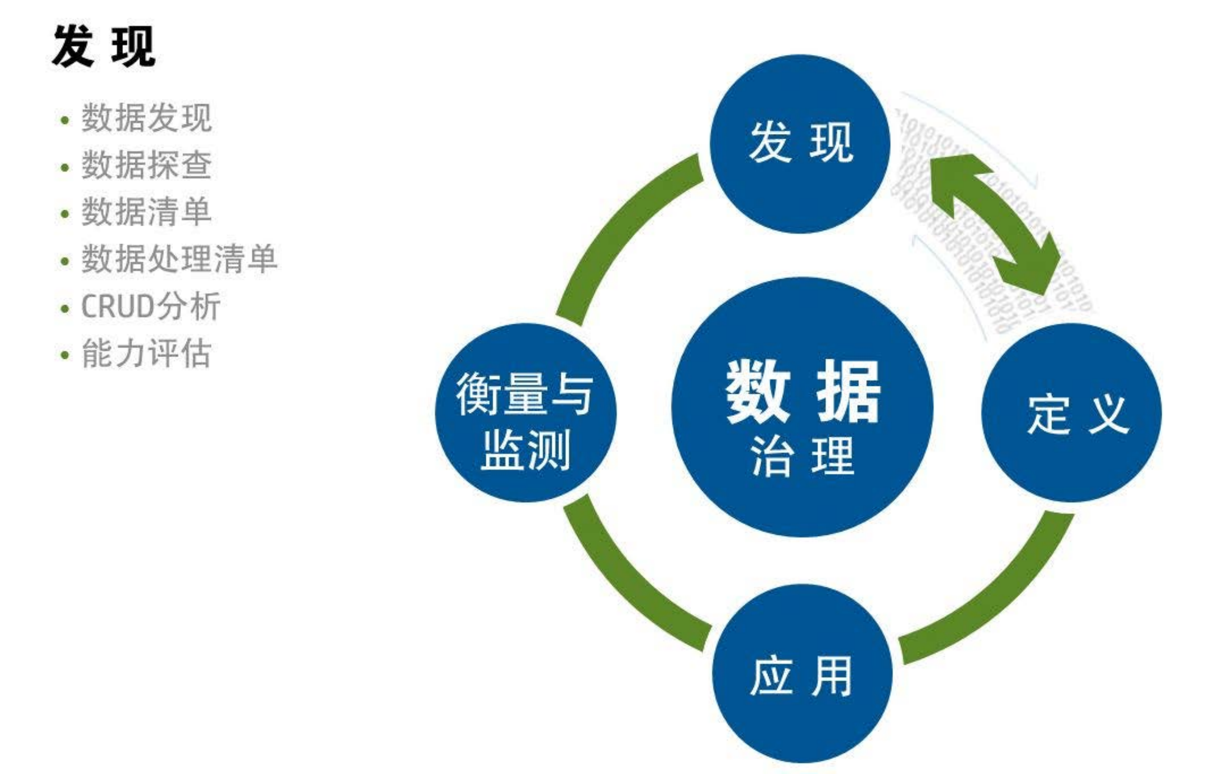
部分日期数据重复报送
Data duplication of several days

从数据采集到最终的数据应用，“数据治理” 占总工作量的约70~80%

高速公路运输量统计
Expressway Passenger and Freight Traffic

从数据采集到最终的数据应用，“数据治理” 占总工作量的约70~80%

From data collection to application， data processing accounts for 70~80% of the total workload



高速公路运输量统计
Expressway Passenger and Freight Traffic

跨省车辆车牌串接算法

跨省的情况下两条记录：

A省：京N12367

B省：京N12361

利用收费站关系、时间、车型、轴数、车重等进行自学习。

The matching of trans-provincial vehicles’ license number

2 records: (1) province A，license number 京N12367 (2) province B，license number 京N12361

The system matches license number according to toll station location, time, vehicle type, vehicle weight and so on

高速公路运输量统计
Expressway Passenger and Freight Traffic

跨省车辆车牌串接算法

车牌号	入境时间	起点		是否起点	途经省市			
		省份名称	地级市名称		省份1	省份2	省份3	省份4
渝[redacted]56	2017-01-13 04:59:33	福建	福州市	是	江西	湖南		
甘[redacted]8	2017-01-13 01:08:43	福建	泉州市	是	江西	湖南		
辽[redacted]02	2017-01-12 22:42:41	河北	秦皇岛市	是	河北	陕西	四川	
冀[redacted]4	2017-01-13 01:03:06	河北	邯郸市	是	河南	四川		
冀[redacted]46	2017-01-12 20:16:02	河北	石家庄市	是	河南	陕西	四川	
苏[redacted]53	2017-01-12 12:49:59	安徽	宿州市	否	河南	湖北	陕西	四川
黑[redacted]5	2017-01-12 13:57:14	黑龙江	哈尔滨市	是	河北	天津	河北	四川
闽[redacted]8	2017-01-13 04:38:21	福建	厦门市	是	江西	湖南		
鲁[redacted]7	2017-01-12 11:26:51	河南	新乡市	是	陕西	四川		
苏[redacted]69	2017-01-12 23:04:31	湖北	十堰市	是	陕西	四川		
渝[redacted]6	2017-01-13 04:55:10	广西	河池市	是	贵州			
吉[redacted]2	2017-01-12 23:58:16	河北	秦皇岛市	否	天津	河北	河南	四川

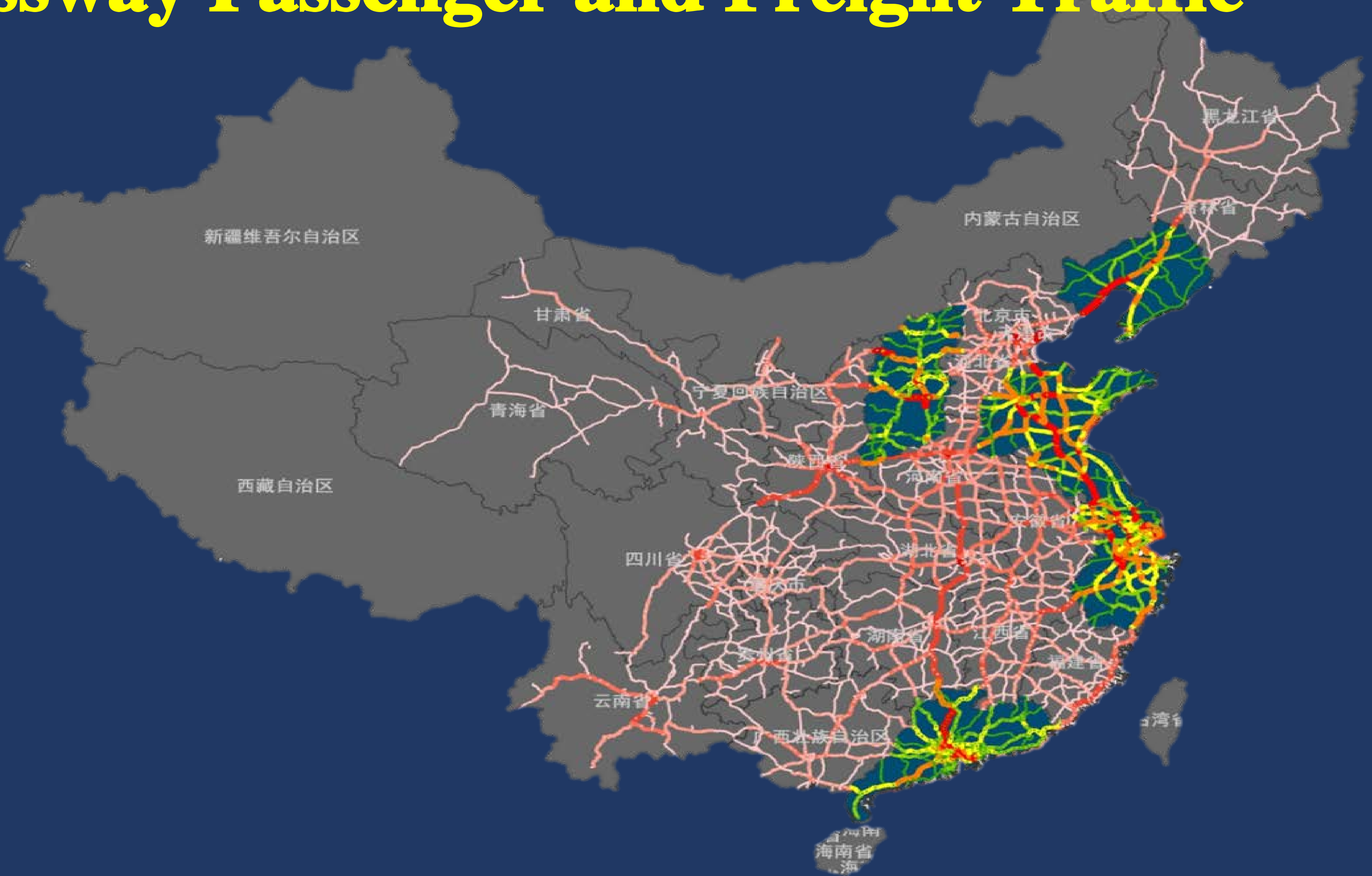
应用机器学习方法，开发基于大数据的跨省车辆串接算法，还原车辆行驶路径。

We developed the matching algorithm of trans-provincial vehicles based on computer learning to track running path

高速公路运输量统计
Expressway Passenger and Freight Traffic

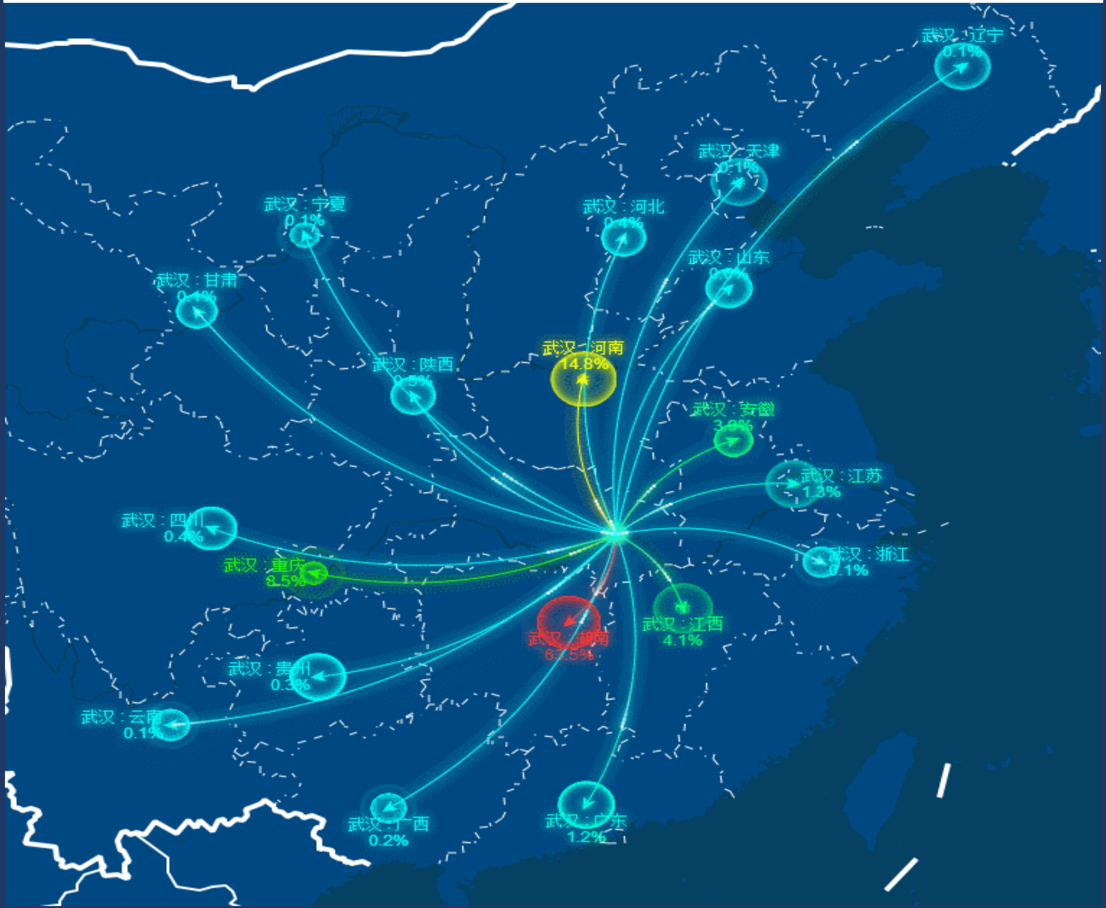
交通流量分配

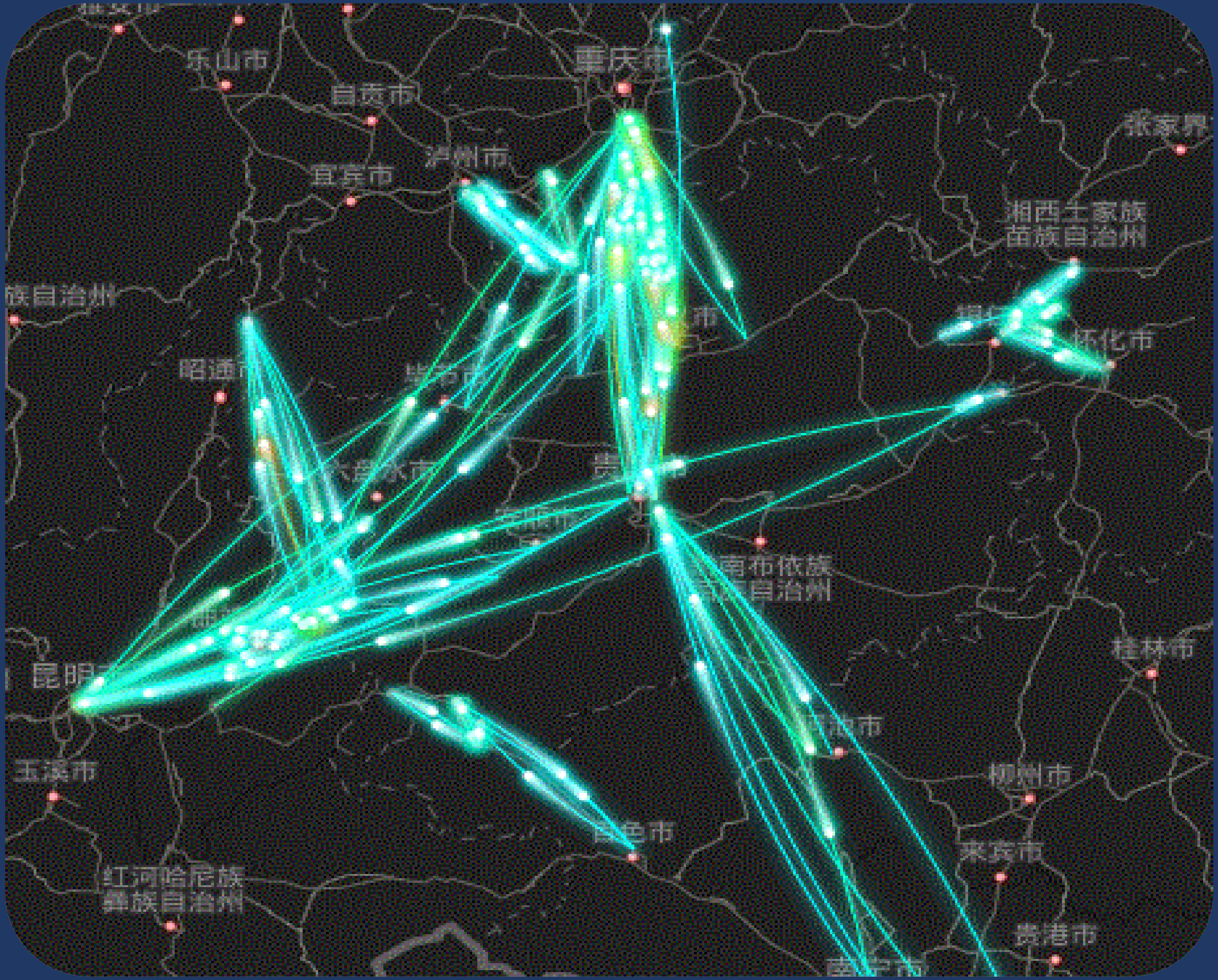
Traffic distribution



高速公路运输量统计
Expressway Passenger and Freight Traffic

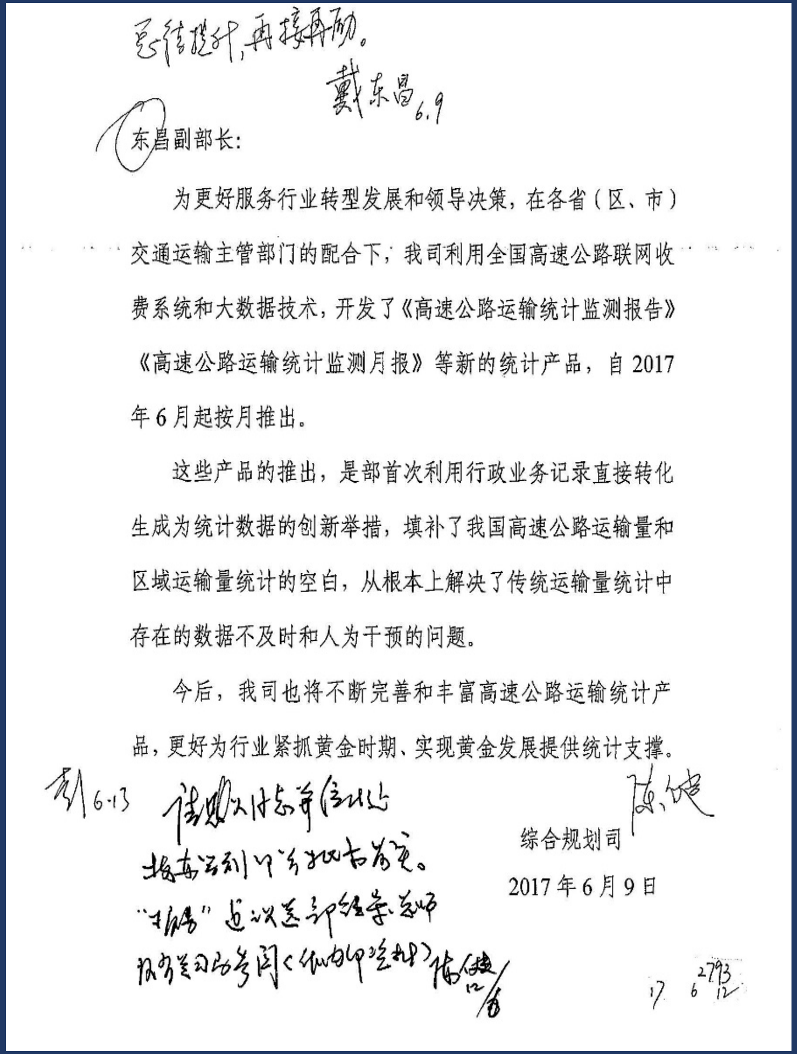
全国高速公路区域OD算法
Expressway O-D algorithm





高速公路运输量统计Expressway Passenger and Freight Traffic

高速公路运输统计监测月报	➡	分区域、分流向、分车型车流量、运输量、周转量、货运密度等19张统计报表，高速公路旅客运输、货物运输情况变化趋势分析
行业统计数据质量评估	➡	以高速公路收费数据中的车货总重、高速公路货运量，结合宏观经济指标，评估并修正各省公路货运量。
高速公路货运量旬报推算	➡	研究确定高速公路货运量旬报推算方法，按旬向国务院报送全国高速公路货运量。
交通运输经济运行分析	➡	月度、季度、年度交通经济运行分析报告。



戴东昌 6.9

东昌副部长：

为更好服务行业转型发展和领导决策，在各省（区、市）交通运输主管部门的配合下，我司利用全国高速公路联网收费系统和大数据技术，开发了《高速公路运输统计监测报告》《高速公路运输统计监测月报》等新的统计产品，自2017年6月起按月推出。

这些产品的推出，是部首次利用行政业务记录直接转化生成为统计数据的创新举措，填补了我国高速公路运输量和区域运输量统计的空白，从根本上解决了传统运输量统计中存在的数据不及时和人为干预的问题。

今后，我司也将不断完善和丰富高速公路运输统计产品，更好为行业紧抓黄金时期、实现黄金发展提供统计支撑。

综合规划司

2017年6月9日

新水路运输统计
Neo-waterway Transport Statistics

内涵contents

利用船舶监管记录信息，包括船舶、船籍港静态信息以及进出港口、装卸货量等动态信息，运用大数据技术进行清洗、处理和挖掘，形成水运量等主要统计指标。

The key data,
such as waterway traffic were calculated by data cleaning, processing and analyzing of big data technology on the basis of vessel administrative records, including vessel information , control country , in-and-out port information, loading and unloading goods information .

新水路运输统计
Neo-waterway Transport Statistics

□覆盖范围：直属海事局、水网地区地方海事局

□港口数量：约400个

□等级航道里程：6.63万公里

□Coverage, main and local branches of MSA

□Port number, 400

□Navigable waterway length up to standard , 66.3 thousand kms

新水路运输统计
Neo-waterway Transport Statistics

□记录条数：海事船舶AIS轨迹数据、进出港口报告数据、进出口岸查验数 据、京杭运河的船闸数据 >400亿条

□数据容量：>10TB（含AIS明细数据）

Records of AIS data, port reporting data, port inspection data and lock data of Jinghang Canal were more than 40 billion

Data capacity, more than 10TB including AIS data



新水路运输统计
Neo-waterway Transport Statistics

□数据内容：进出港口或口岸的船舶静态信息（船名、各类登记号、船舶尺度、船舶吨位、船籍、经营人）动态信息（船舶位置、装卸客货量、货物种类等）

□Data contents—vessel name, license number, vessel size, vessel tonnage, control country, management entity, vessel location, passenger and freight traffic, freight categories

新水路运输统计
Neo-waterway Transport Statistics

应用服务层：主要是通过BI、报表、查询等工具，直接面向行业决策、咨询、研发等需求。

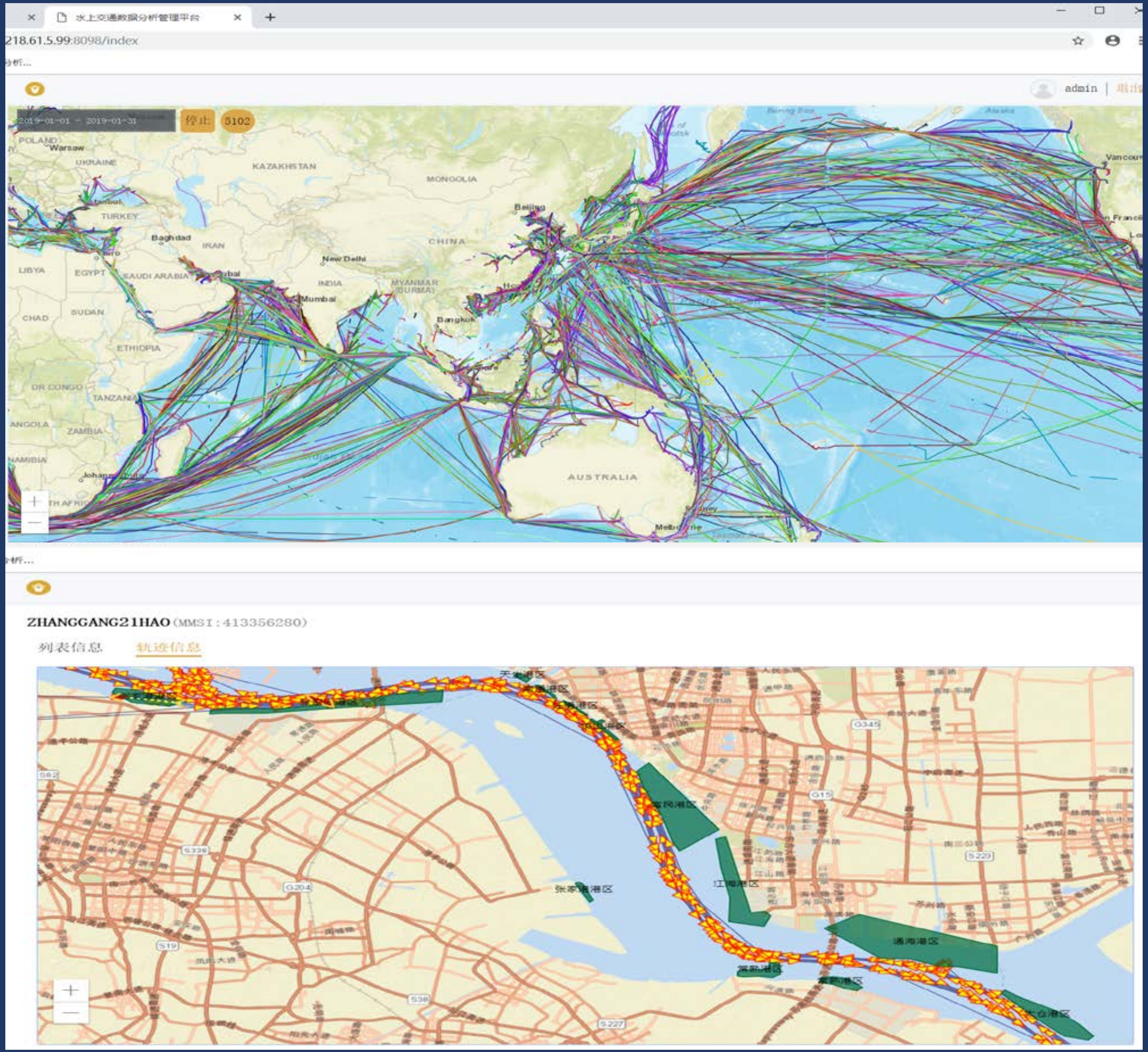
支撑技术层：对数据进行清洗、挖掘、分析的核心层，通过关键技术的支撑来保障数据质量、实现多源数据融合计算核心统计指标。

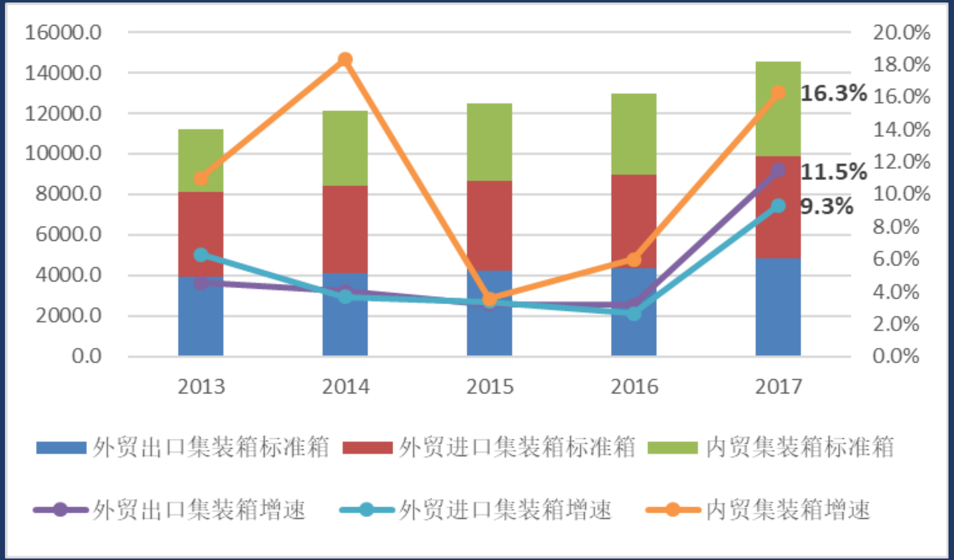
基础设施层：采用基于Hadoop体系对航运大数据进行存储、解析、流式处理，进行特征提取。

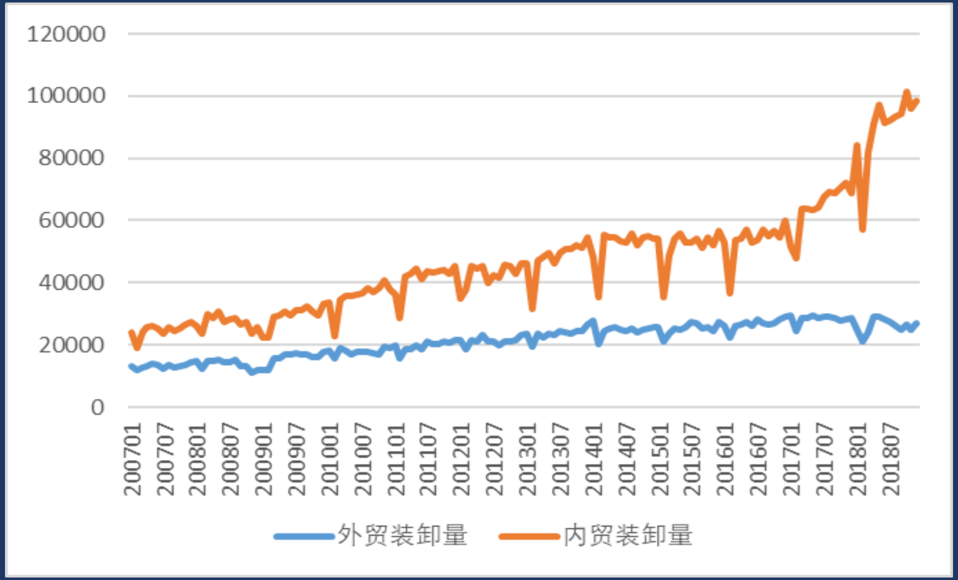
数据交换层：依托于部级共享交换平台。

数据采集层：依托行政相对人和市场主体等行业行政监管记录的采集主体，将数据采集延伸至移动网络所及之处，实现了全国范围水运数据的在线采集。

新水路运输统计
Neo-waterway Transport Statistics







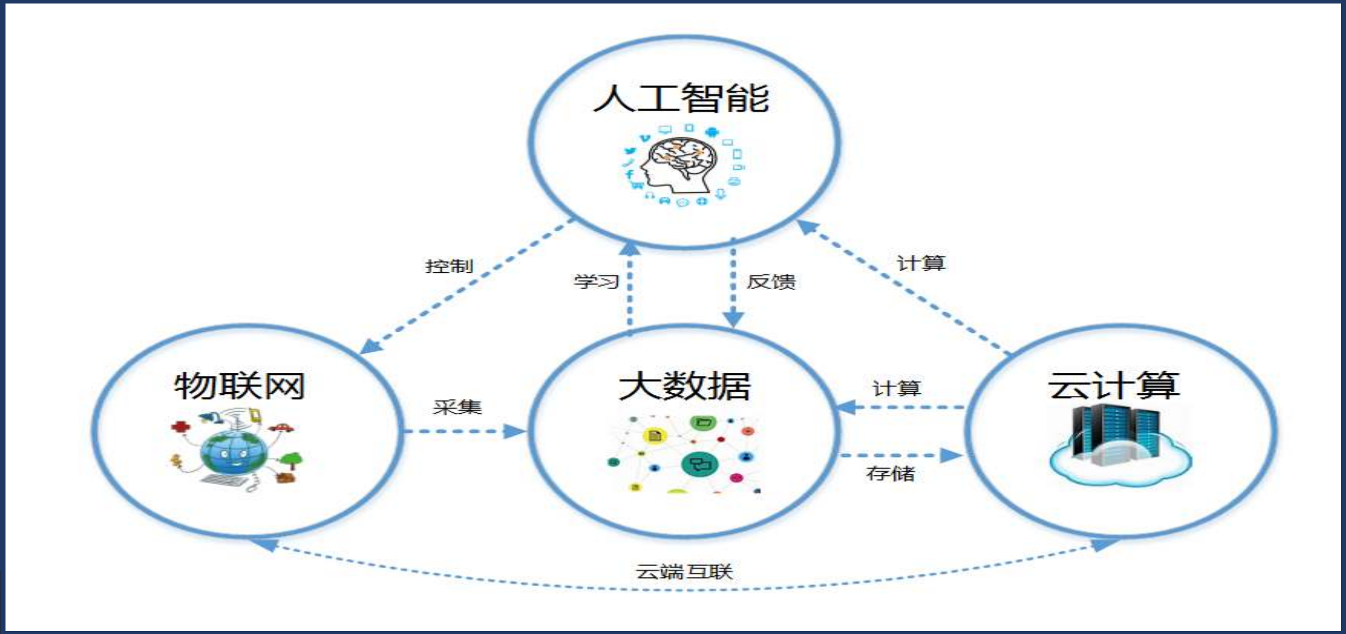
四点体会和感受

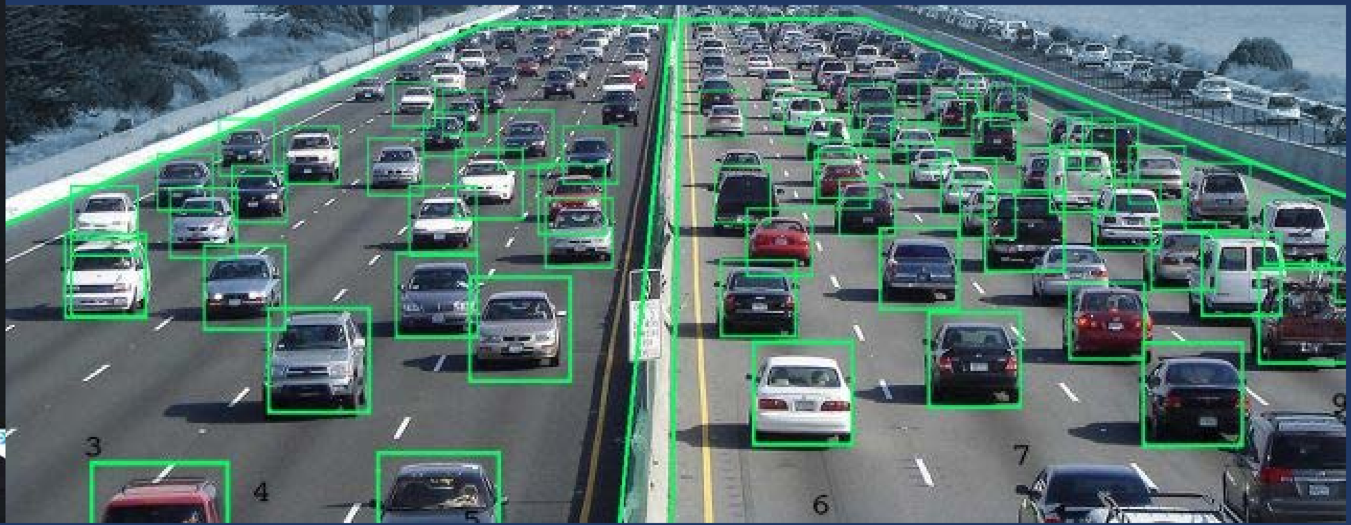
Experiences and feelings

方向与决心不容动摇
Determination and Goals

思想与理念必须革新
Innovative thinking

技术与方法抓紧完善
Techniques and Methods Improvement





体制与机制亟需建立
Institutions and mechanisms are urgently needed

共同努力，多方发力，

推进大数据统计应用开花结果！

Working together to create favorable outcomes of Big Data application in transportation statistics