



挖掘多源数据 发现城市拥堵规律

IBM 王君



CONTENTS

01 多源数据融合思路

02 拥堵模型方案设计

03 工程实践结果展现



01

多源数据融合思路

- ◆ 数据：丰富但多源
- ◆ 问题：如何最大化利用数据？
- ◆ 挑战：在数据准备的每个环节
- ◆ 思路：按需所取，优势互补

交通运营数据

卡口, RFID, 车联网, 轨道闸机
GPS, 手机信令

1

交通设施数据

道路、轨道、公交、停车场、公路、铁路、
港口、机场、步行与自行车, 地下管网, 加
油气站等

2

交通规划数据

交通规划(总规, 控规, 综合, 专项, 研
究), 规划用地、现状用地、现状建筑、建
筑平面图

3

交通地理数据

电子政务地图、卫星(航拍)影像图、地形
图(1:500)、地名数据

4

样本量
GPS样本偏少
车联网样本丰富



覆盖范围
RFID路网分布间隙大
GPS不覆盖隧道
卡口路网分布密度大

缺失值
各数据源有不同程度的缺失值



数据互补
提高样本量，减少缺失值



统一数据
建立一套基础数据指标体系

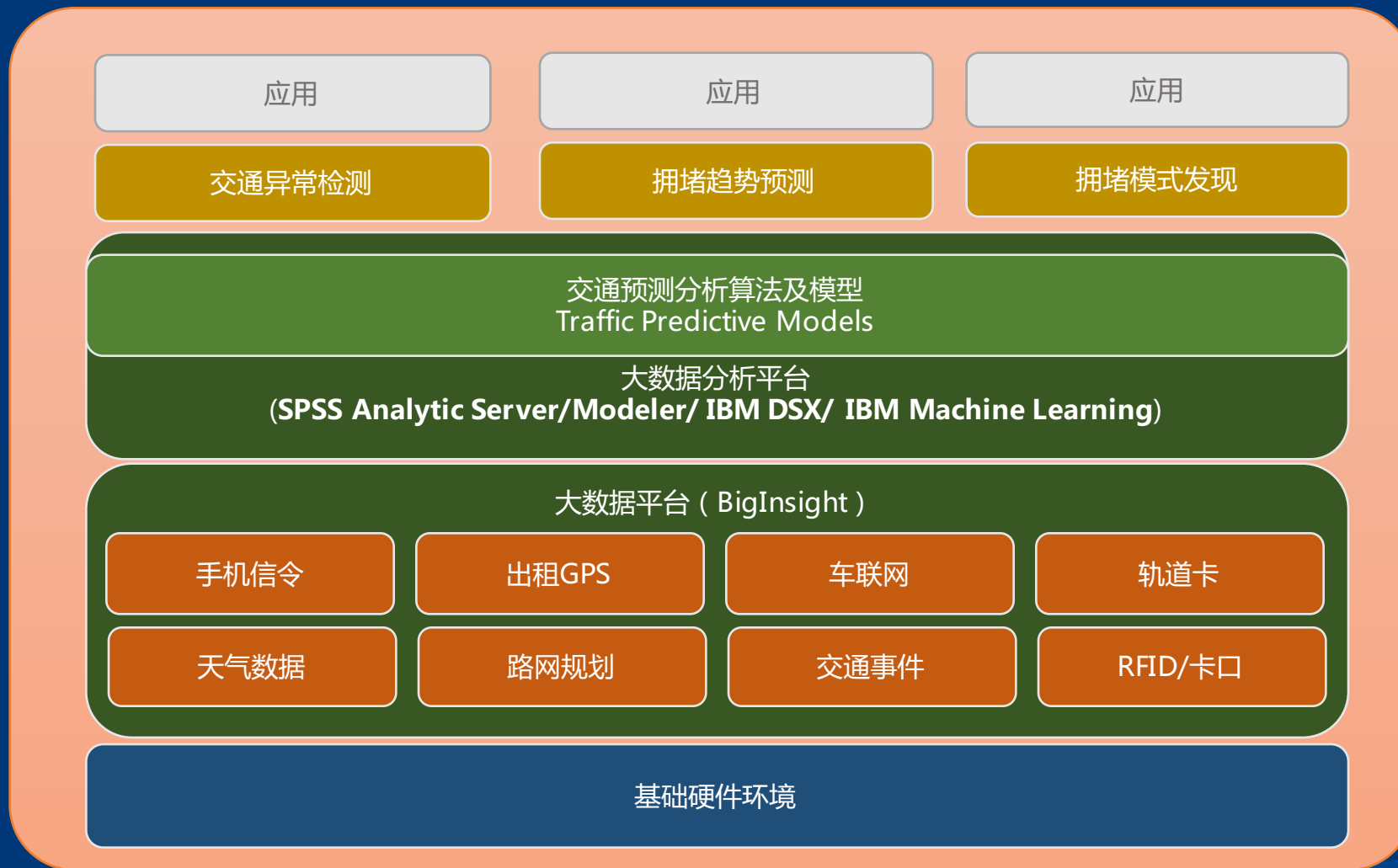




02

拥堵模型方案设计

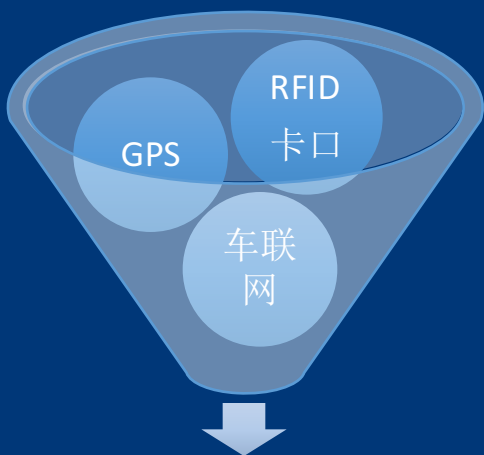
- ◆ 方案架构
- ◆ 方案内容
- ◆ 方案价值



方案内容

1. 路段指标形成

- 从原始数据中抽象出能代表路段交通状态的各类指标如：
 - 车速
 - 低速行驶里程
 - 低速行驶时间
 - ...



交通指标体系

2. 拥堵事件的自动识别

- 拥堵事件是识别常发拥堵路段的基础
- 全新的、智能的方法，自动识别交通路网中所有的拥堵事件，为后续分析奠定可靠的基础

3. 拥堵事件的刻画

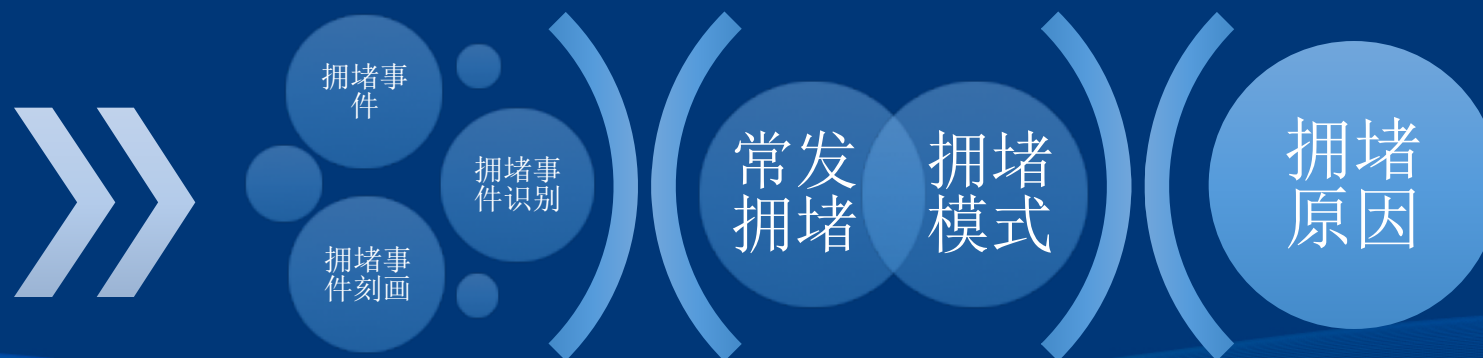
- 拥堵事件的刻画是对拥堵事件进行特征性描述
- 对拥堵事件准确的刻画是拥堵路段识别的数据来源。

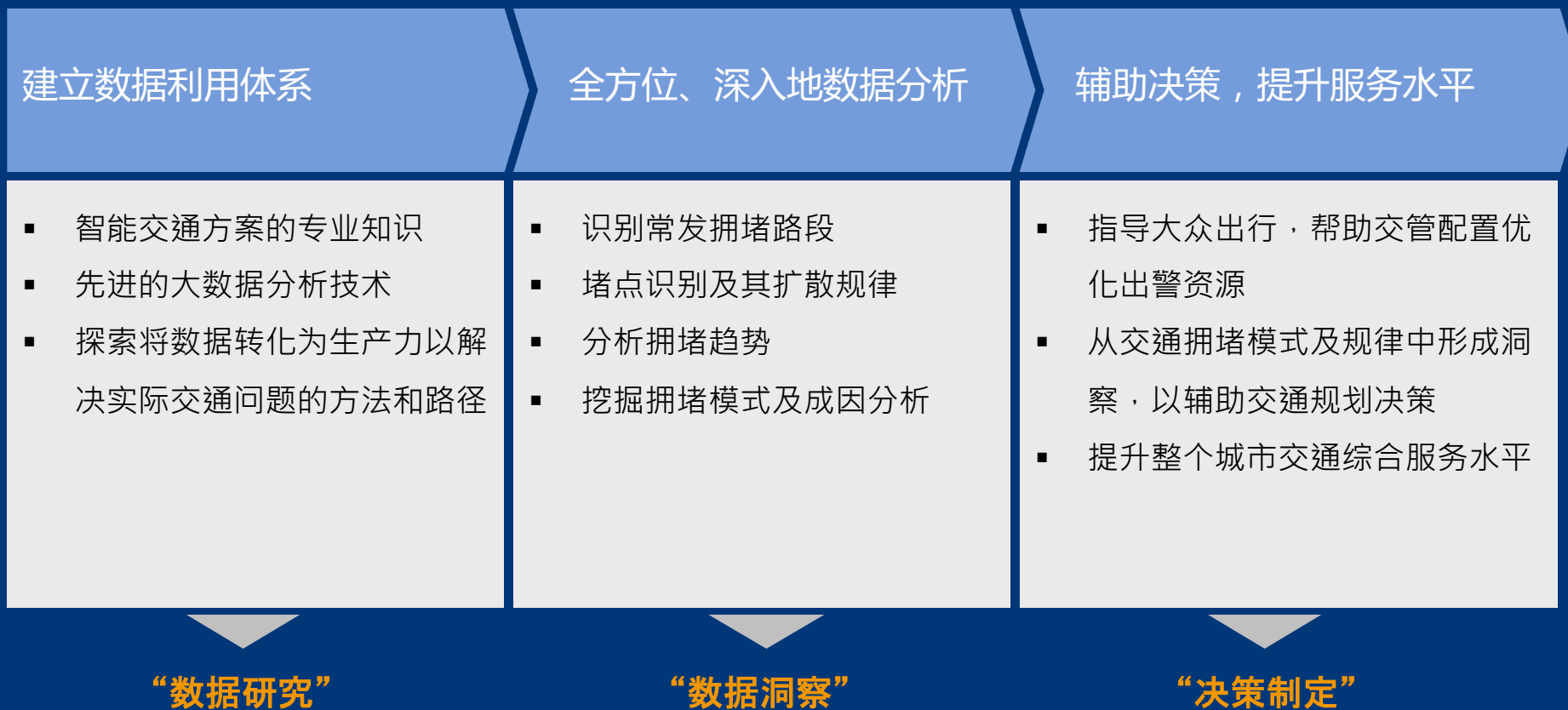
4. 常发拥堵路段识别

- 对常发拥堵的路段进行识别，能够更好地定位拥堵核心点，辅助决策者进行缓堵决策

5. 拥堵模式分析

- 在现实道路交通流中，路段拥堵模式均不相同，不同模式的拥堵对整个交通网络的影响也不同。
- 对不同拥堵模式下，常发拥堵路段的识别，对拥堵原因提供了更深入的见解







03

工程实践结果展示

- ◆ 模型实现
- ◆ 开发原则
- ◆ 结果展示



模型设计

模型可行性确定

- 模型算法选择
- 基本参数确定

初步验证基于

- 一个月数据
- 小路网
- 小型Cluster



模型验证

模型验证环境搭建

模型迁移及验证

- 在更大数据集上
- 区域路网
- 中大型Cluster

性能测试及调优

基于验证结果模型更新及重构

模型工程化

- 健壮性
- 稳定性



模型实施

生产环境集成与测试

全数据测试

集群性能调优

算法模型

准确性
可扩展性
健壮性



迭代，快速迭代

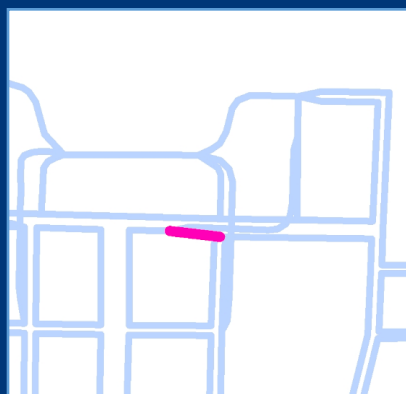
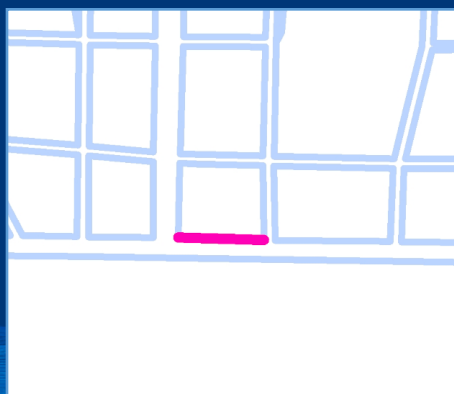
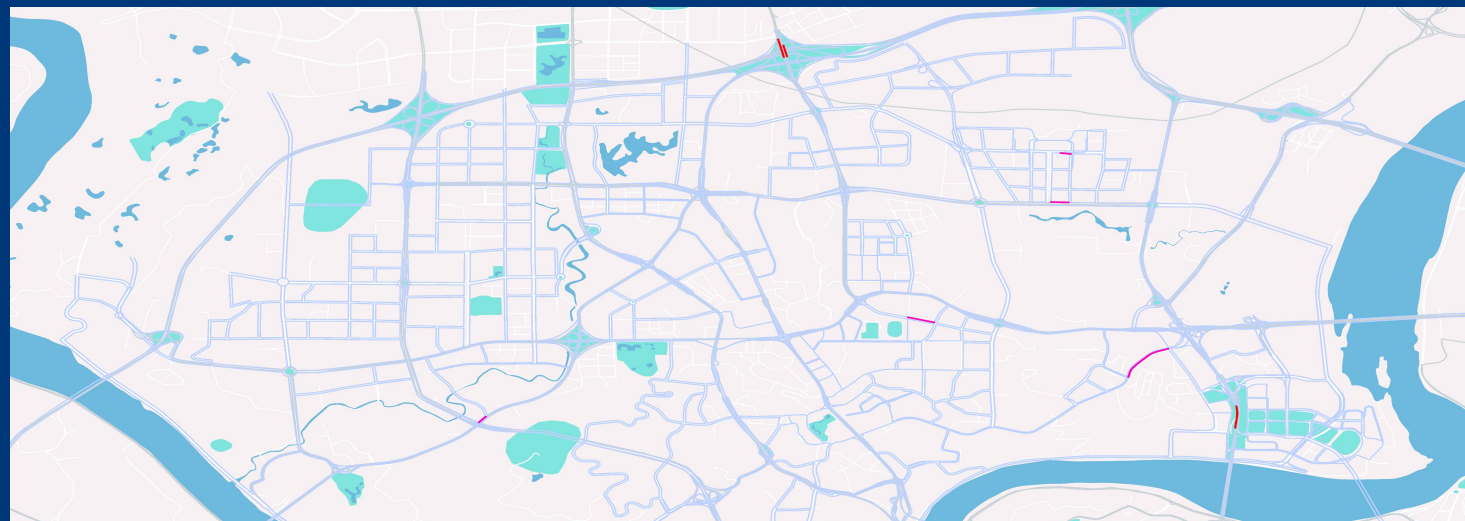
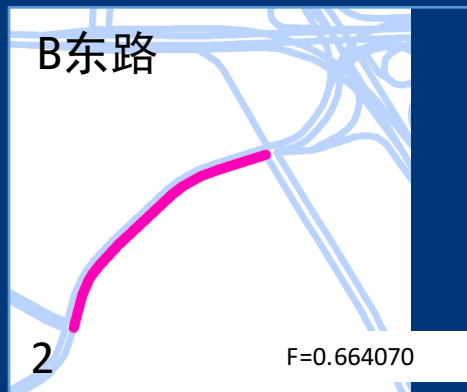
细分功能场景
逐个快速迭代

应用平台

低耦合
集群参数优化

拥堵常发路段

7条常发拥堵路段



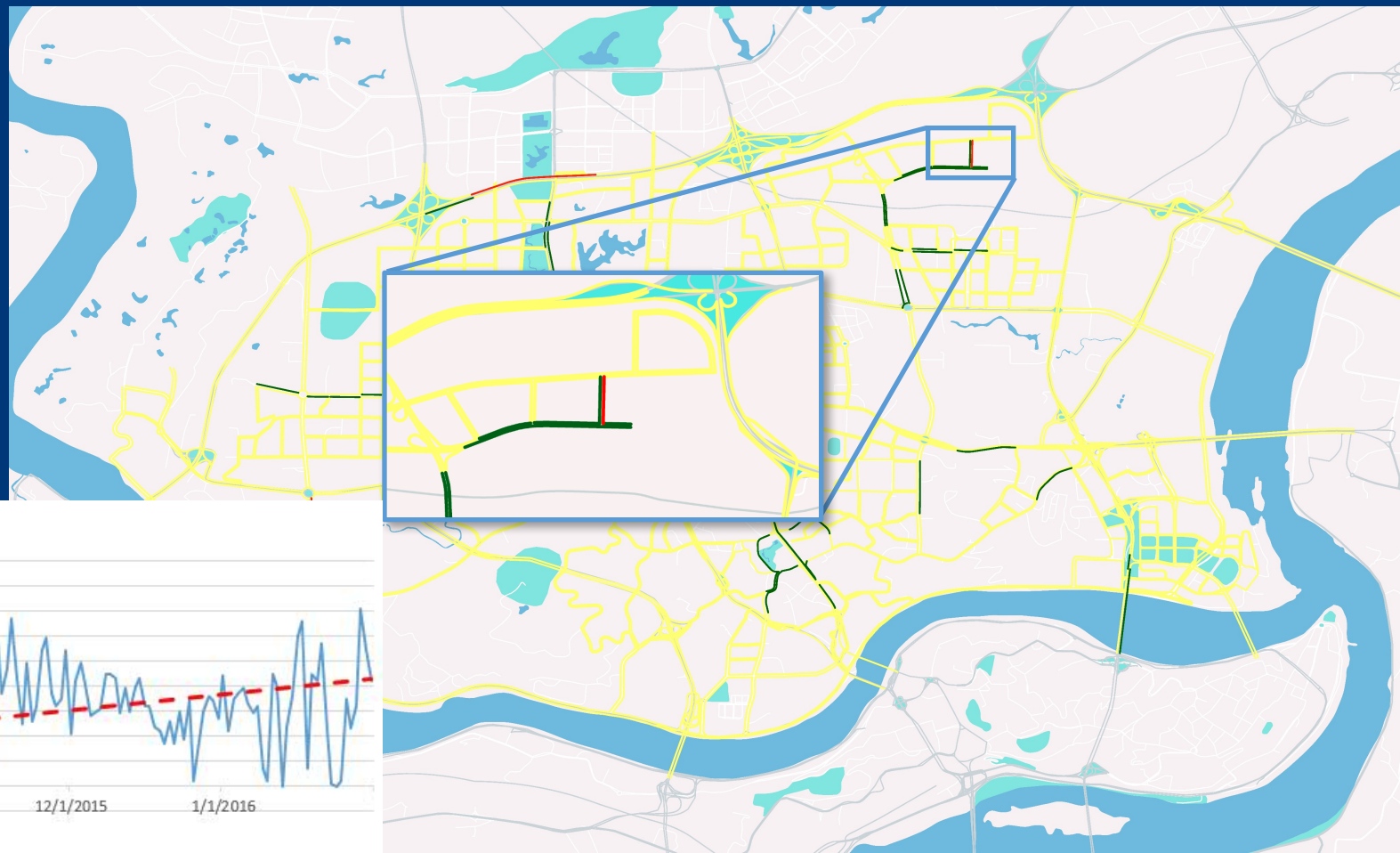
拥堵趋势变化 – 整体情况

通过7个月历史数据探索路段拥堵总体趋势，得到以下结论

- 432条主干道中
 - 0条路段逐渐向好
 - 4条路段拥堵情况持续恶化
- 176条快速路中
 - 9条路段拥堵情况逐渐向好
 - 0条路段持续恶化
- 52条高速路中
 - 1条路段拥堵情况逐渐向好
 - 1条路段拥堵情况持续恶化

拥堵趋势变化- 示例图

- 恶化最严重的主干道路段
- EE大道南段（由南向北）



谢谢

2017
中国·上海



IBM资深大数据架构师：王君



13319262429



wjuncdl@cn.ibm.com



IBM中国研发实验室