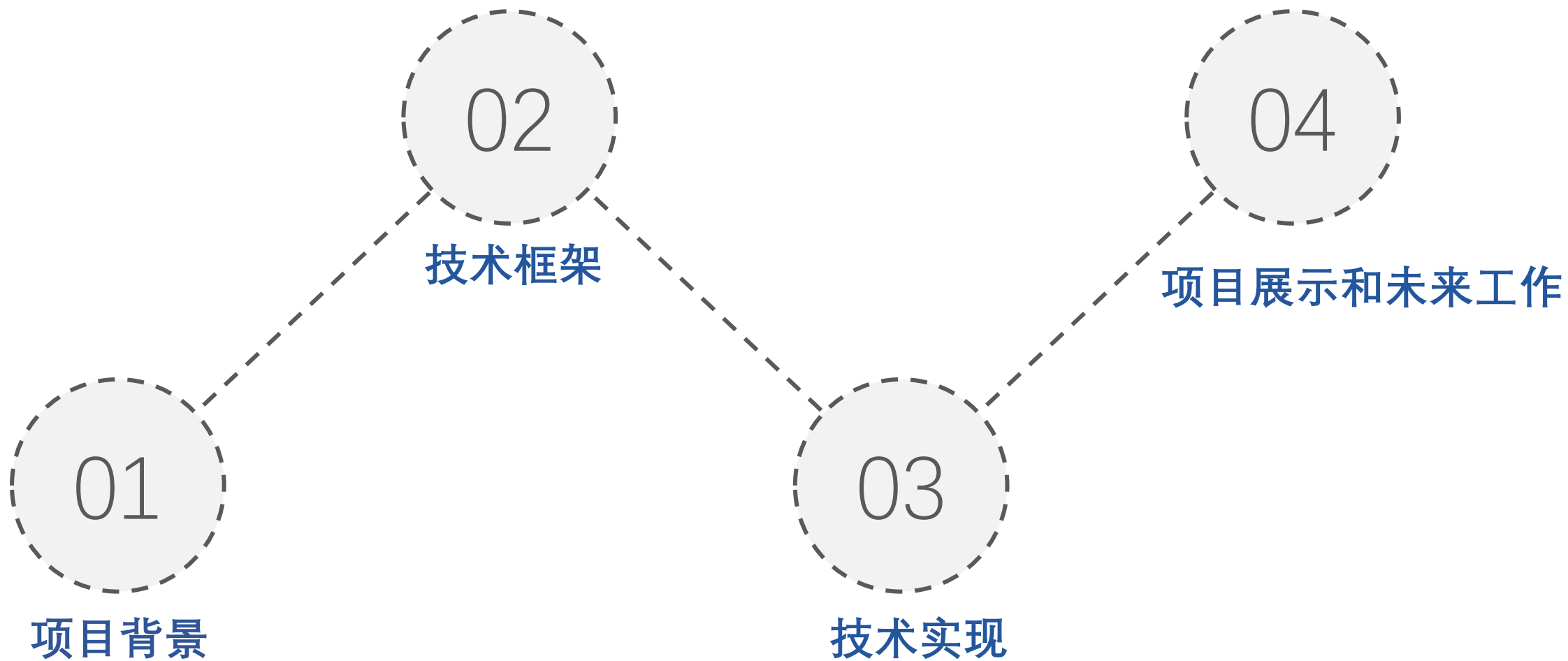




太原理工大学
TAIYUAN UNIVERSITY OF TECHNOLOGY



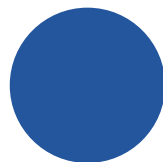
基于大数据和强化学习的 旅游路线推荐系统



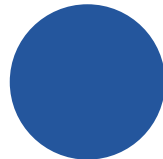


项目背景

01



旅游路线优化需求



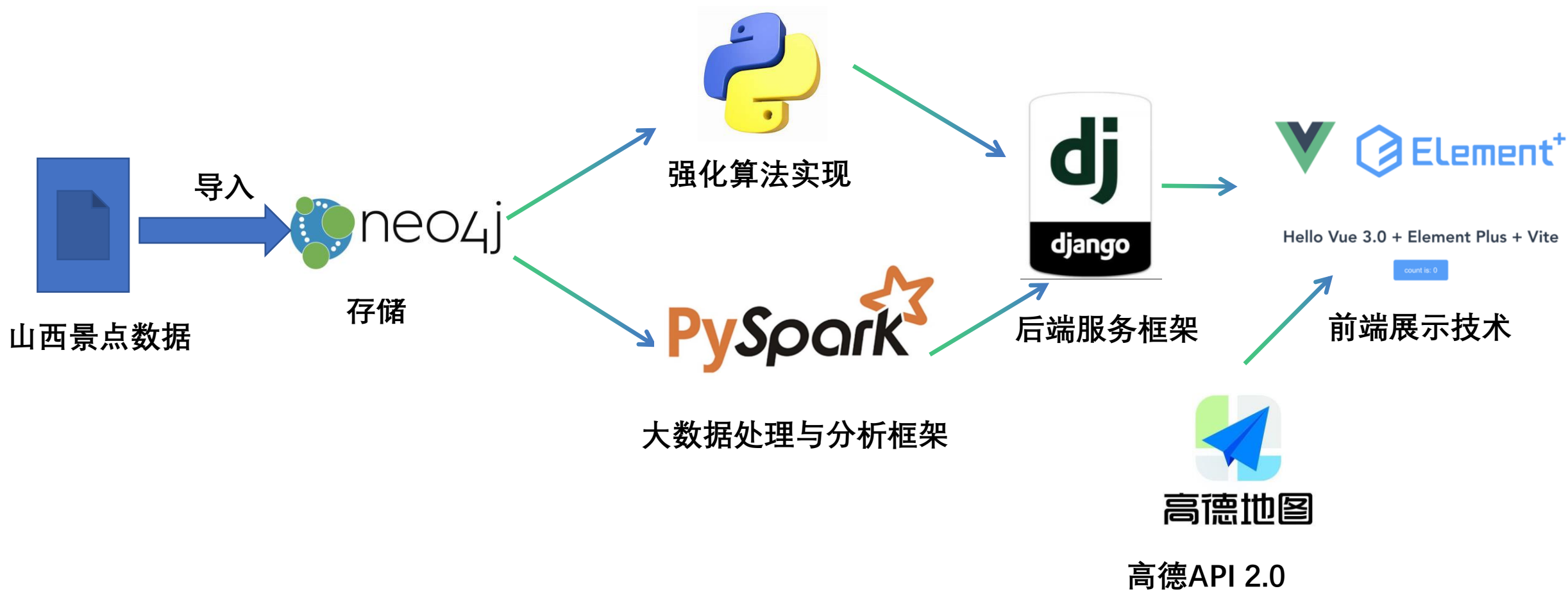
技术路线

旅游路线优化需求

随着旅游业的发展，自驾游作为一种**自由、灵活**的旅游方式，越来越受到游客的青睐。尤其在山西这样的历史文化大省，丰富的自然景观和人文遗址为自驾游提供了广阔的空间。然而，面对**众多的景点和复杂的路线选择**，游客往往难以制定出既符合自身兴趣又高效的游览计划。

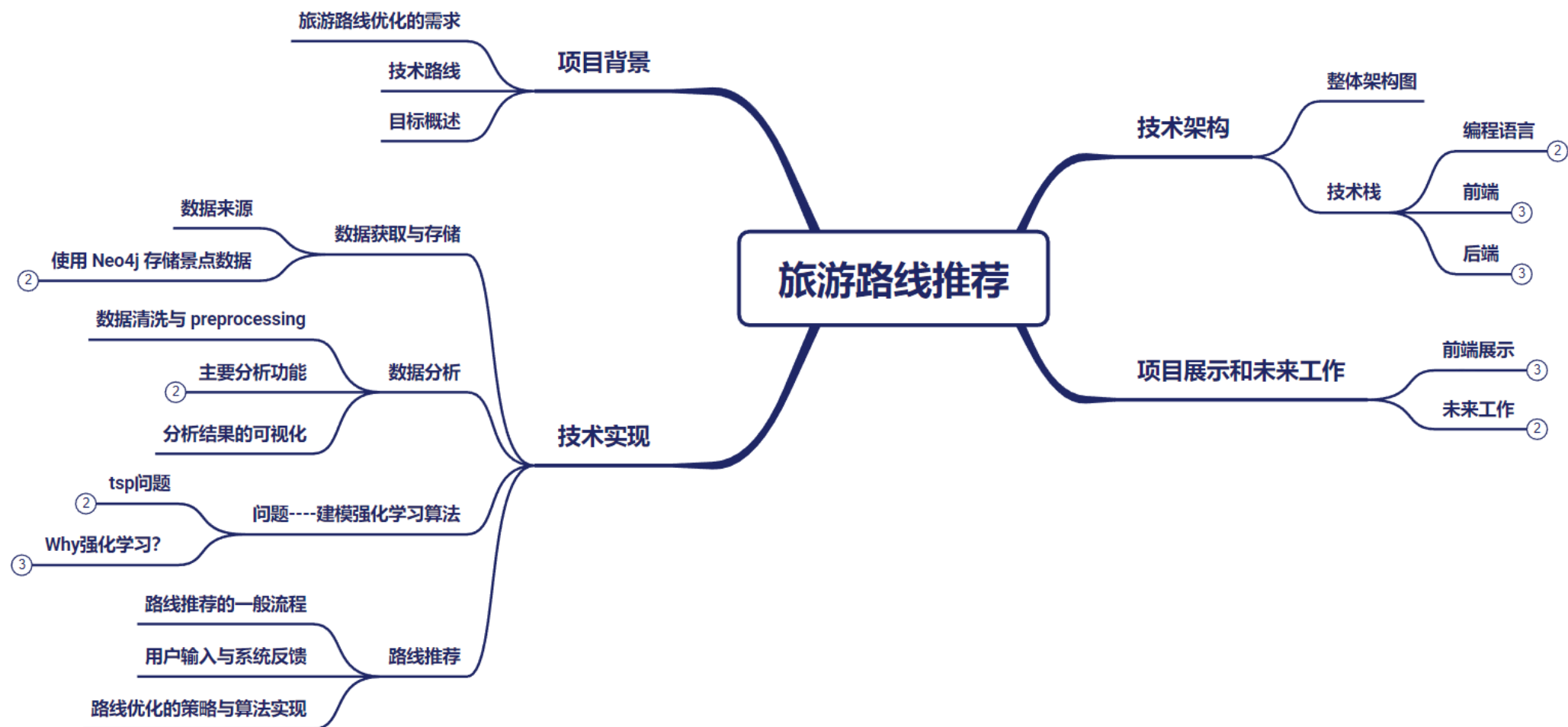


技术路线





思维导图





技术架构

02

1

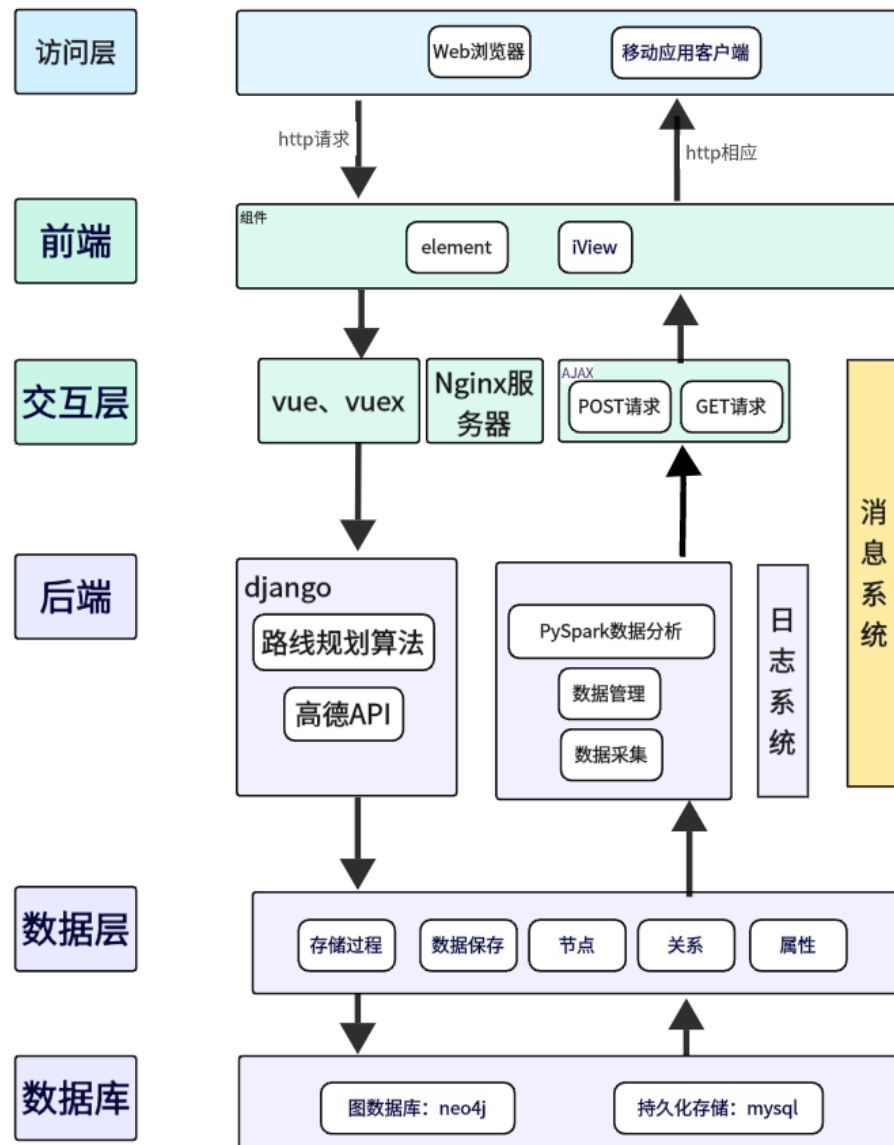
系统架构

2

技术栈介绍

系统架构

- Web访问
- 前端接受http请求
- Ngnix反向代理均衡负载
- 强化学习路径规划
- Pyspark数据分析
- 大型数据存储



FRONT-END

(一) 高德 API 2.0

新特性预览



多边形吸附



MultiPolygon



LabelMarker与主图标注避让



zoom范围扩展至 [2,20]

- Canvas 图层 [CanvasLayer](#)
- 任意切片图层 [TileLayer.Flexible](#)
- 完全自定义图层 [CustomLayer](#)
- 热力图 [Heatmap](#)

地图控件

- 比例尺 [Scale](#)
- 缩放工具条 [ToolBar](#)
- 鹰眼 [HawkEye](#)
- 图层切换 [MapType](#)
- 定位按钮 [Geolocation](#)

信息窗体

- 信息窗体 [InfoWindow](#)

右键菜单

- 右键菜单 [ContextMenu](#)



高性能矢量标注 LabelMarker

- 椭圆 [Ellipse](#)
- 矩形 [RectAngle](#)
- 弧线 [BezierCurve](#)

坐标变换

坐标系变换

- 经纬度 <-> 地图容器坐标
- 经纬度 <-> 地图平面坐标

地图坐标转换

- GPS(WGS84) -> 高德(GCJ02、火星)
- 百度(Baidu) -> 高德
- 图吧(MapBar) -> 高德

轨迹动画

- 定位
- 初始定位 [Map](#)
- 城市定位 [CitySearch](#)
- 浏览器定位 [Geolocation](#)

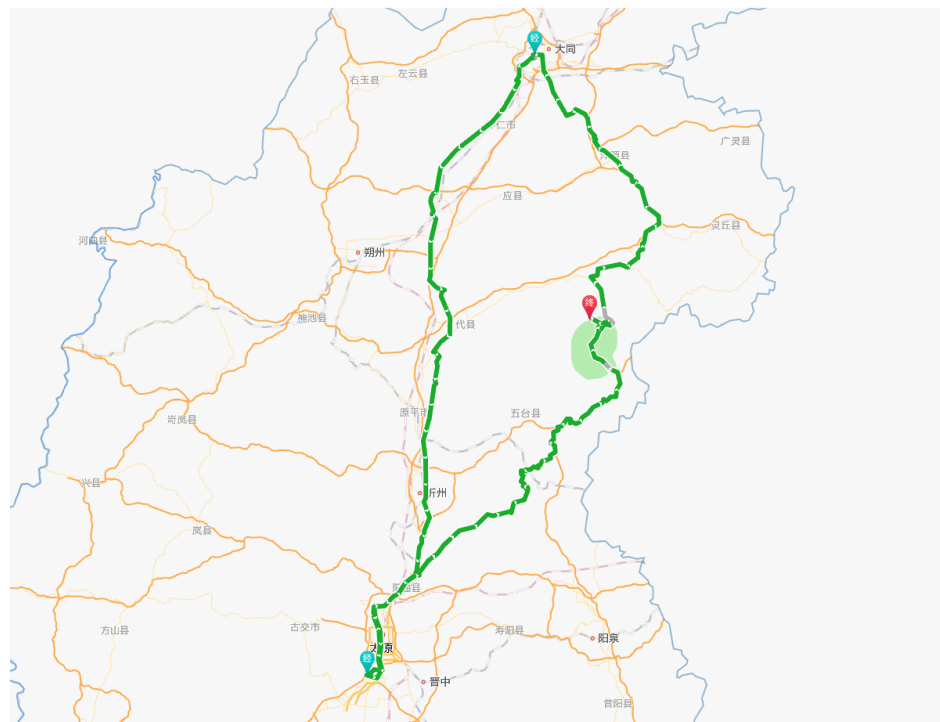
其他服务

- 天气 [Weather](#)
- 公交站点 [StationSearch](#)
- 公交线路 [LineSearch](#)

支持库

- 空间几何计算 [GeometryUtil](#)
- 通用函数库 [Util](#)
- DOM函数库 [DomUtil](#)
- 浏览器信息 [Browser](#)

地图 JS API 2.0以WebGL为主要绘图手段

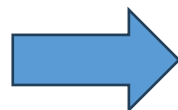
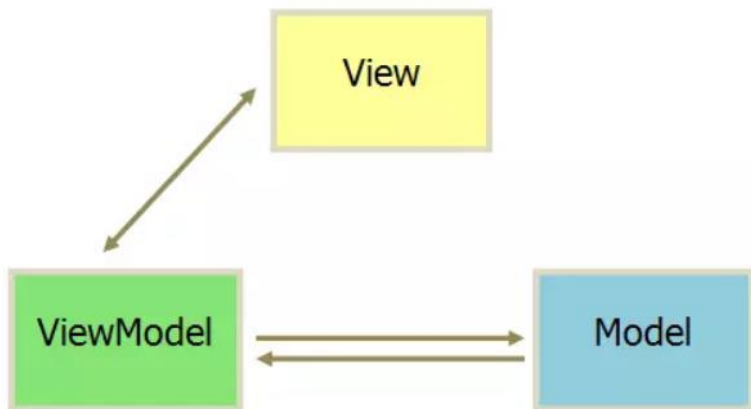


FRONT-END

(二) Vue、Element Plus

构建用户界面的渐进式框架 (MVVM 模式)

- Model 层代表数据模型
- View 代表UI 组件
- ViewModel 是一个同步View 和 Model的对象



Element Plus 是一个基于 Vue 2.0 的桌面端组件库（超链接、按钮、图片、表格、选项等）




The image shows a web interface for travel route planning. It includes a sidebar with filters and a main content area with route details and a list of stops.

旅行策略

最快捷模式

起点/终点

五台山

途经点

- ☐ 五台山
- ☒ 云冈石窟
- ☒ 晋祠博物馆
- ☐ 应县木塔
- ☐ 山西黄河壶口瀑布
- ☐ 平遥古城

搜索路线

具体路线信息

预计行驶时间 **950 分钟**

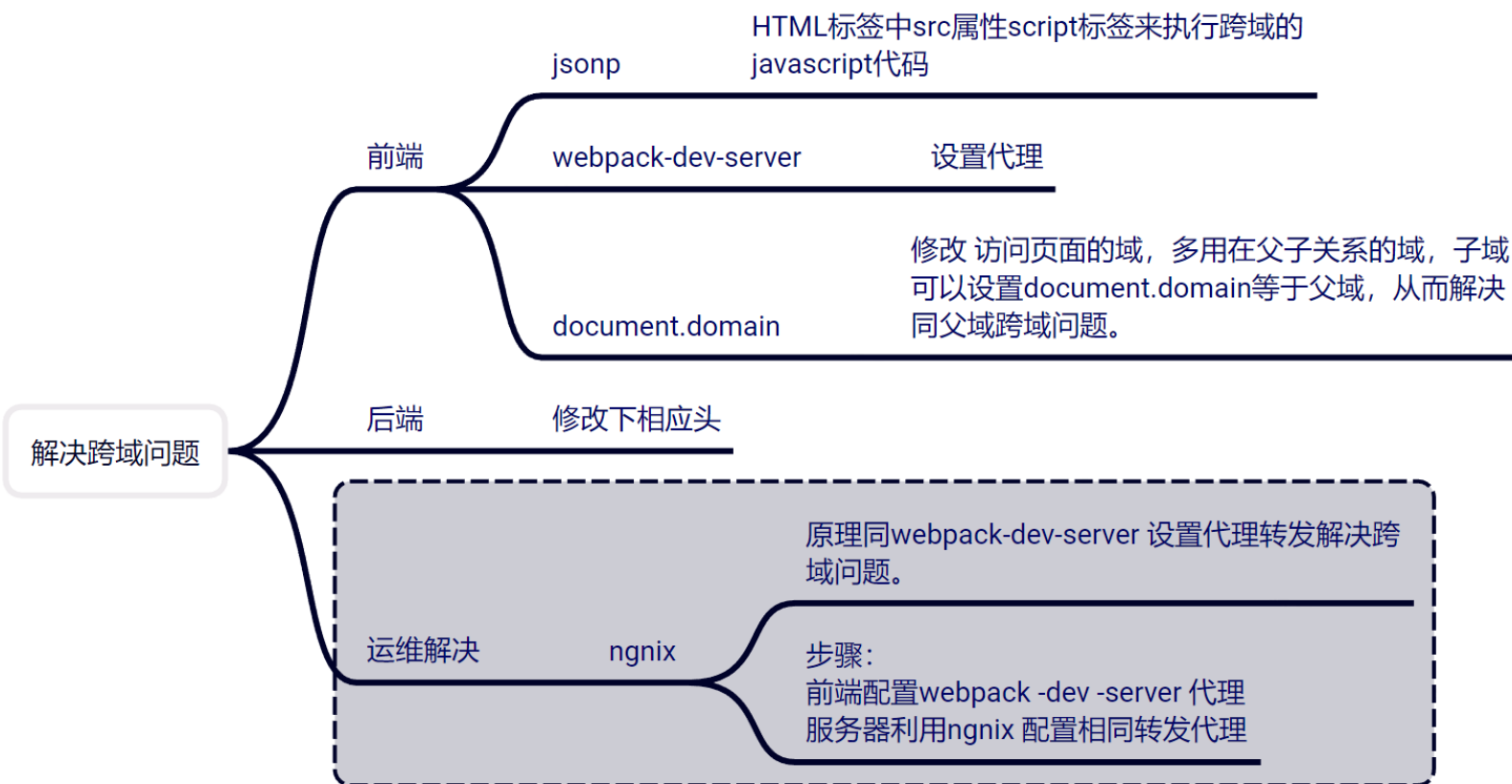
路线距离 **796.1 公里**

路径规划

- 晋祠博物馆
- 云冈石窟
- 五台山
- 晋祠博物馆

FRONT-END

(三) nginx负载均衡、跨域反向代理



- 跨域：浏览器受同源策略（协议相同、域名相同、端口相同）的限制
- 不允许不同源的站点之间进行某些操作（如发送Ajax请求）。
- 通过进行一些特殊配置是解决→使用 **nginx** 完成跨域反向代理

BACK-END Neo4j

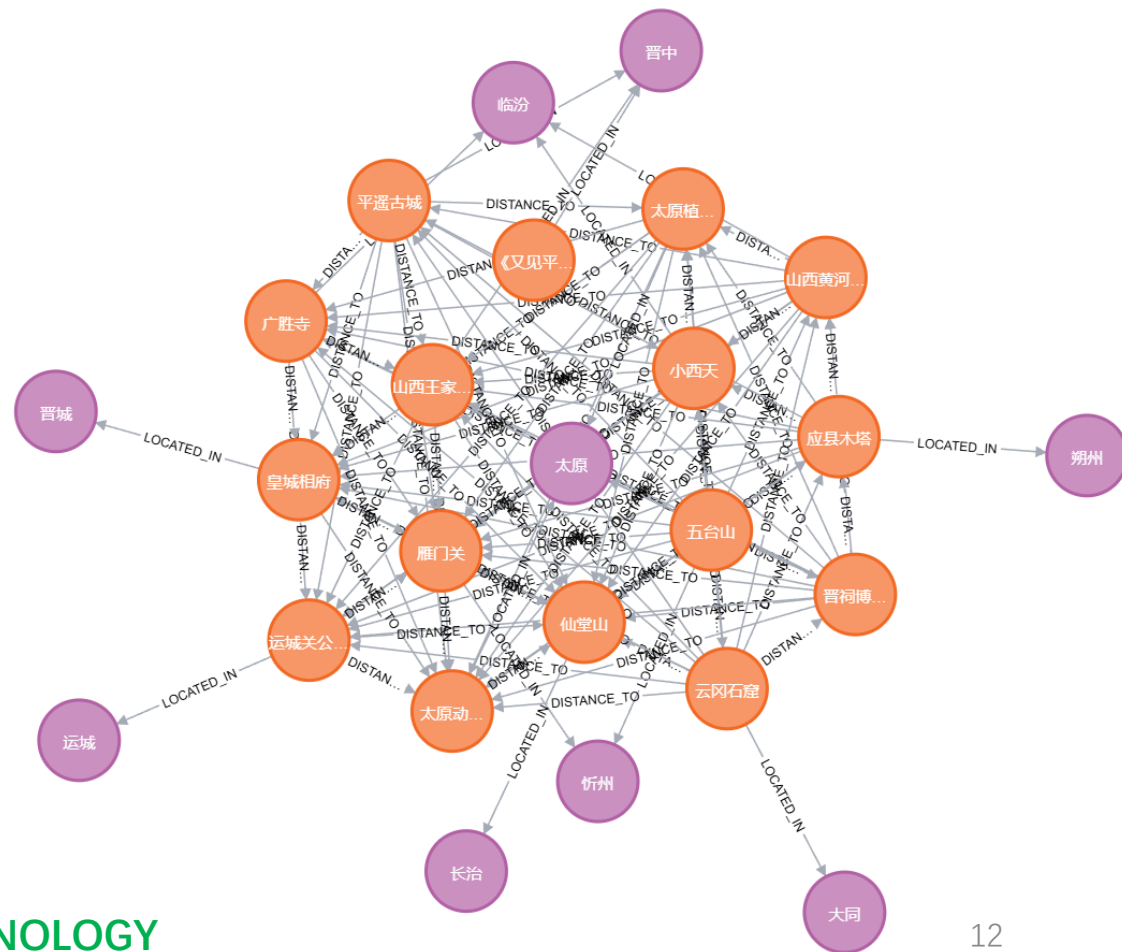
基于图论实现的一种NoSQL数据库，其数据存储结构和数据查询方式都是以图论为基础的，图数据库主要用于存储更多的连接数据。

图数据库数据模型的主要构建模块是：

- 节点
- 关系
- 属性

标签（Label）将一个公共名称与一组节点或关系相关联

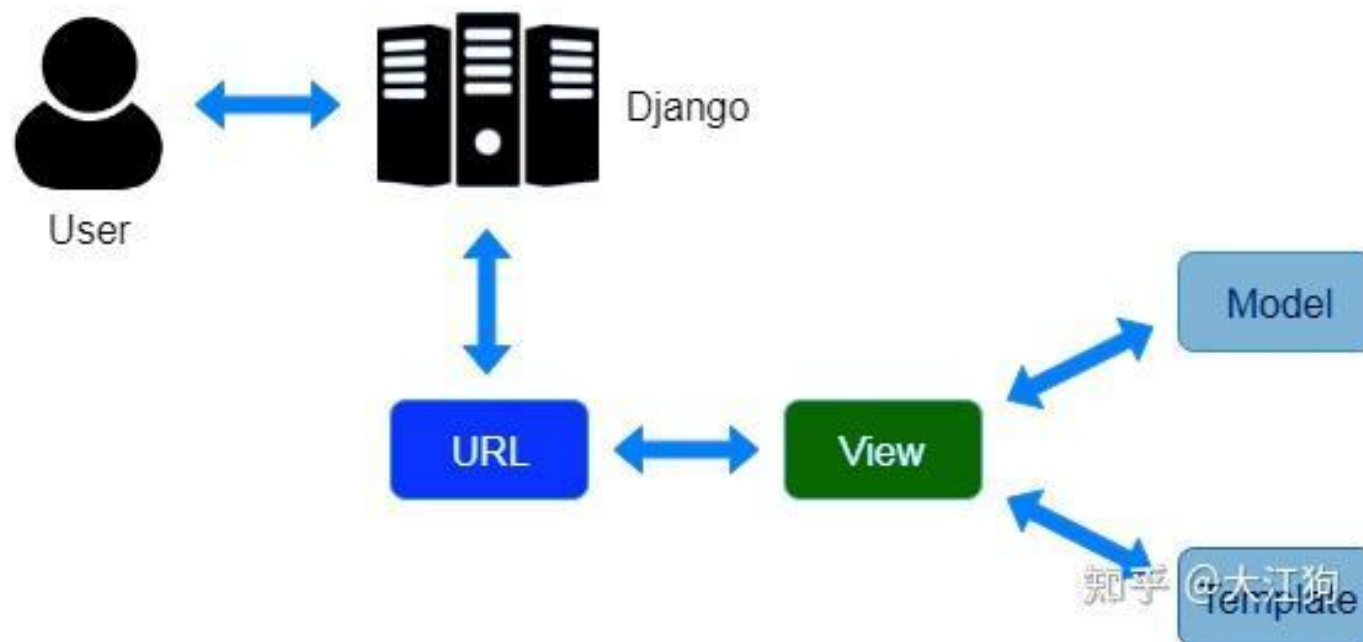
- 一个节点可以具有多个标签
- 对标签进行索引以加速在图中查找节点



BACK-END Django

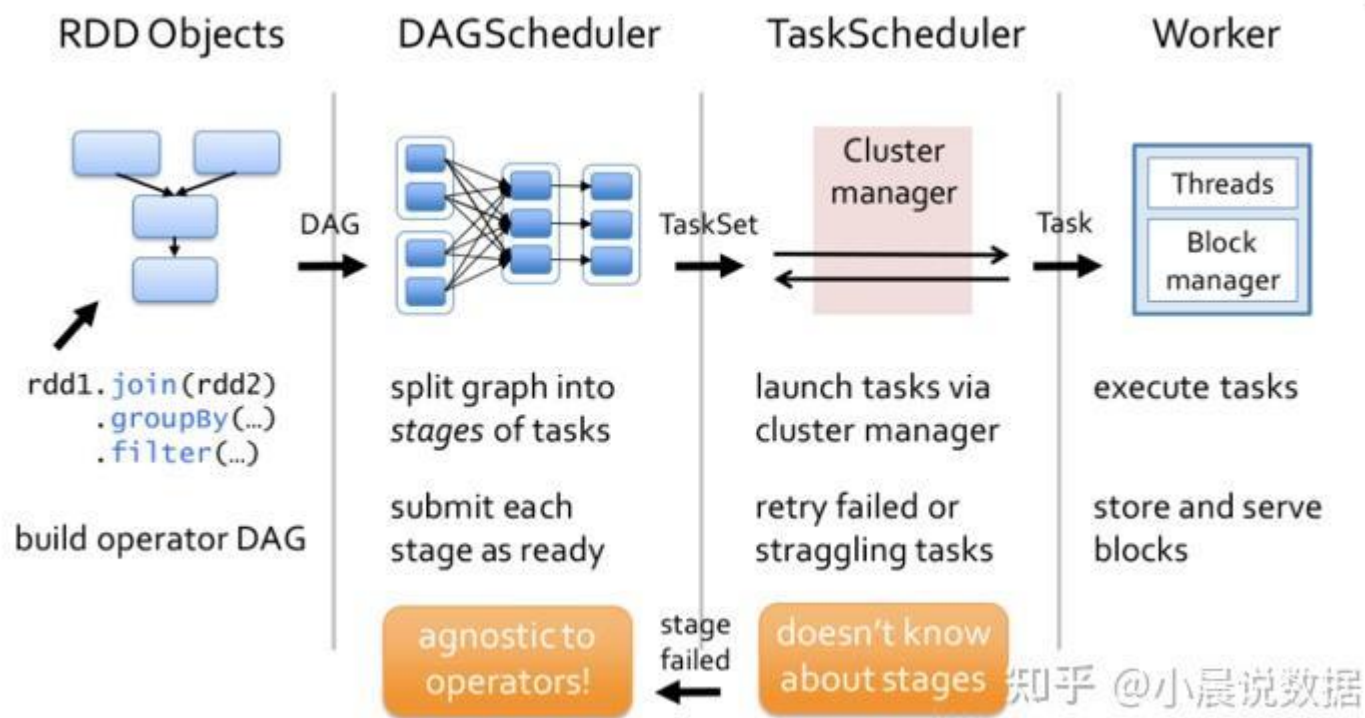
构建 Web 应用程序的高级 Python Web 框架。

采用了 MVT 的软件设计模式，即模型（Model），视图（View）和模板（Template）



BACK-END PySpark

基于Python的Spark编程接口，可以用于大规模数据处理、机器学习和图形处理等各种场景。Spark是一个开源的大数据处理框架，它提供了一种[高效的](#)分布式计算方式。



知乎 @小晨说数据



技术实现

03

- 数据获取与存储
- 数据分析
- 问题建模---强化学习算法
- 路线推荐

(一) 数据获取——景点基本信息数据来源

数据来源

- 去哪儿旅行官网

输入山西，筛选景点

数据获取

- BeautifulSoup爬虫技术
- 节点、标签筛选
- Page循环爬取



(一) 数据获取——景点基本信息数据信息

数据信息

- 景区名称
- 星级
- 地区
- 地址
- 评分
- 价格
- 简介
- 详情页
- 营业额

数据预处理

- 数据清洗：删除多余的标点符号
- 格式统一

	A	B	C	D	E	F	G	H	I	J
1	景区	星级	地区	地址	评分	价格	简介	详情页		
2	五台山	5A景区	山西·忻州	山西省忻州	4.9	115	中国佛教四大名山之首	https://piao.qunar.com/ticket/det		
3	云冈石窟	5A景区	山西·大同	山西省大同	5	100	中国三大石窟之一	https://piao.qunar.com/ticket/det		
4	晋祠博物馆	4A景区	山西·太原	山西省太原	5	80	晋国宗祠园林建筑群	https://piao.qunar.com/ticket/det		
5	应县木塔	4A景区	山西·朔州	山西省朔州	5	48	辽代古建筑，由木头建成	https://piao.qunar.com/ticket/det		
6	山西黄河壶	5A景区	山西·临汾	山西省临汾	4.9	90	看滚滚黄河水呼啸而下	https://piao.qunar.com/ticket/det		
7	平遥古城	5A景区	山西·晋中	山西省晋中	5	23	晚清时期的中国金融中心	https://piao.qunar.com/ticket/det		
8	小西天	4A景区	山西·临汾	山西省临汾	5	32.8	欣赏精美的明代彩塑	https://piao.qunar.com/ticket/det		
9	太原植物园	4A景区	山西·太原	山西省太原	5	30	太原植物园是市委、市政府确定的重	https://piao.qunar.com/ticket/det		
10	广胜寺	4A景区	山西·临汾	山西省临汾	5	68	全国首批重点文物保护单位	https://piao.qunar.com/ticket/det		
11	山西王家大	4A景区	山西·晋中	晋中市灵石	4.9	30	清代民居建筑集大成者	https://piao.qunar.com/ticket/det		
12	皇城相府	5A景区	山西·晋城	山西省晋城	4.7	73.4	明清城堡式官宦住宅建筑群	https://piao.qunar.com/ticket/det		
13	《又见平遥》大型实景		山西·晋中	晋中市平遥	4.8	238	讲述了古老票号的家族故事	https://piao.qunar.com/ticket/det		
14	运城关公庙	4A景区	山西·运城	运城市盐湖	4.9	45	关羽家乡的关帝爷祖庙	https://piao.qunar.com/ticket/det		
15	雁门关	5A景区	山西·忻州	山西省忻州	5	53.6	依山势而建的长城古关道	https://piao.qunar.com/ticket/det		
16	太原动物园	4A景区	山西·太原	太原市杏花	5	10	观赏众多珍稀动物	https://piao.qunar.com/ticket/det		
17	仙堂山	4A景区	山西·长治	山西省长治	4.6	60	青山叠翠、奇峰峥嵘	https://piao.qunar.com/ticket/det		
18	广仁王庙		山西·运城	山西省运城	4.3	13.5	又称五龙庙，在芮城县城北面四公里	https://piao.qunar.com/ticket/det		
19	贾家庄民俗动物园		山西·吕梁	吕梁市汾阳	5	9	参观民俗馆，欣赏动植物	https://piao.qunar.com/ticket/det		
20	龙山石窟		山西·太原	山西省太原	4.7	7	国内少见的道教石窟	https://piao.qunar.com/ticket/det		
21	汤如意城市汤泉生活馆		山西·太原	太原市小店	4.6	96		https://piao.qunar.com/ticket/det		
22	红岩岭自然风景区		山西·阳泉	山西省阳泉	4.8	19.9	溶洞多且壮观，秋赏红叶	https://piao.qunar.com/ticket/det		
23	蒲津渡遗址		山西·运城	山西省运城	4.9		渡口铁牛	https://piao.qunar.com/ticket/det		
24	名邸温泉		山西·吕梁	吕梁市孝义	4.7		孝义名邸温泉酒店是由山西名企金晖	https://piao.qunar.com/ticket/det		
25	华门	4A景区	山西·临汾	山西省临汾	4.7	35	纪念帝尧统一华夏的功绩	https://piao.qunar.com/ticket/det		
26	池神庙		山西·运城	山西省运城	5	39.9	歇山式屋顶，雕梁画栋，精致壮观	https://piao.qunar.com/ticket/det		

(一) 数据获取——景点距离和位置数据来源

景点距离数据来源

调用高德地图的api获取到各景点之间距离的数据

	A	B	C	D
1	出发景点	目的景点	驾车距离（公里）	
2	五台山	云冈石窟	199.99	
3	五台山	晋祠博物馆	258.45	
4	五台山	应县木塔	210.41	
5	五台山	山西黄河壶	495.61	
6	五台山	平遥古城	338.25	
7	五台山	小西天	466.62	
8	五台山	太原植物园	258.45	
9	五台山	广胜寺	462.4	
10	五台山	山西王家大	393.1	
11	五台山	皇城相府	580.68	
12	五台山	运城关公故	624.03	
13	五台山	雁门关	166.34	
14	五台山	太原动物园	231.57	

景点位置数据来源

调用高德地图api获取经度、纬度信息

景点名	景点地址	经度	纬度
五台山	山西·忻州·五台县	113.5647	39.0792
云冈石窟	山西·大同·南郊区	113.298	40.07525
晋祠博物馆	山西·太原·晋源区	112.4779	37.71552
应县木塔	山西·朔州·应县	112.4329	39.33173
山西黄河壶口瀑布旅游区	山西·临汾·吉县	111.5193	36.08858

(一) 数据存储

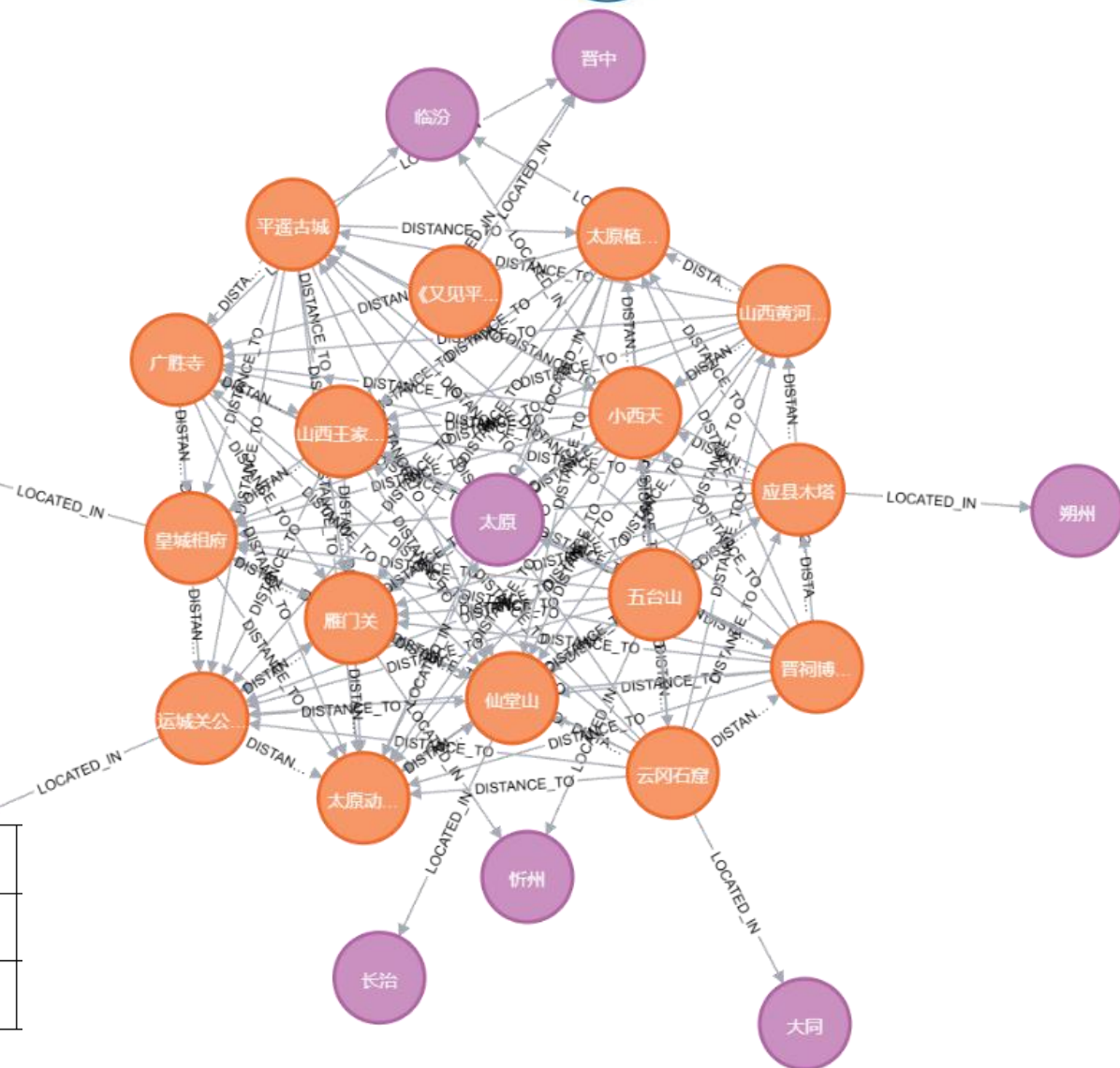
Neo4j数据库存储数据信息

节点:各个景点

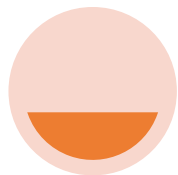
属性: 各个景点的属性信息, 包括星级, 地址等信息

边: 景点之间的距离; 景点位于某地区

关系	含义	节点	含义
LOCATION_IN	景点位于某地区	ScenicSpot	景区名称
DISTANCE_TO	景点间距离	Location	市级地区名

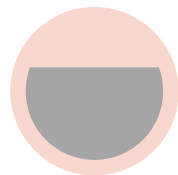


(二) 数据分析



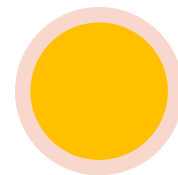
数据清洗与处理

- 删除数据中多余的标点符号
- 根据数据分析需求, 通过计算平均值, 统计获取需要的数据
- 处理技术使用 pyspark



主要分析功能

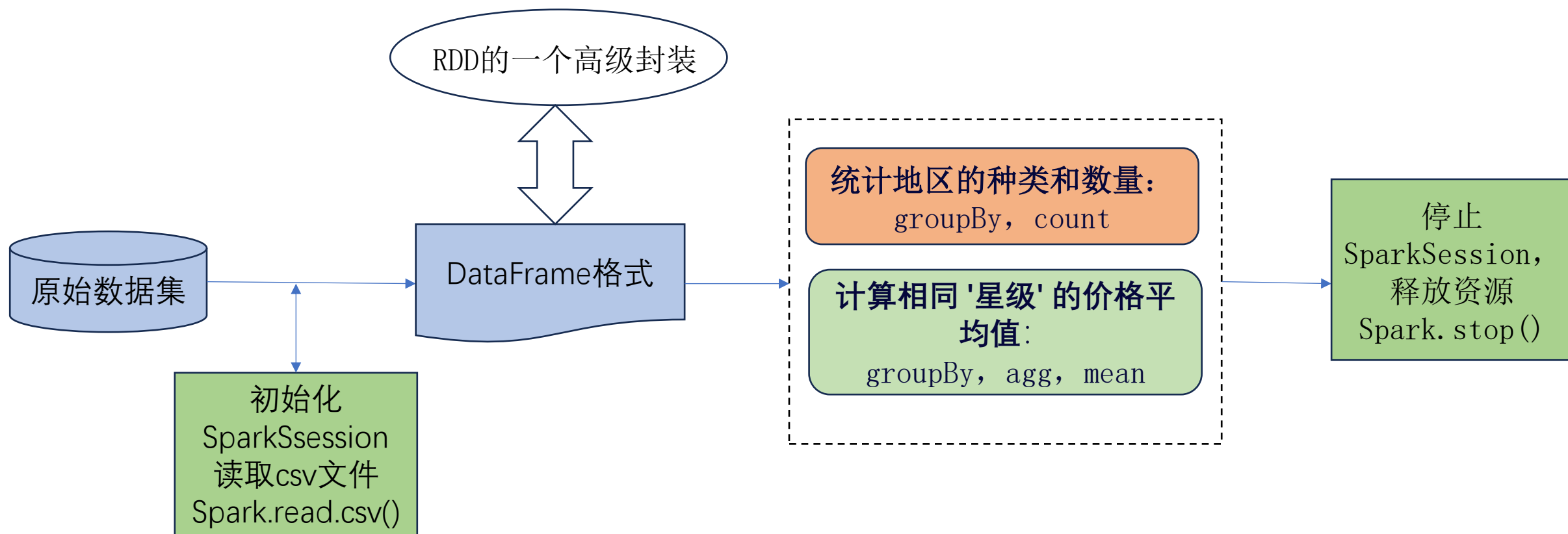
- 景点分布城市占比
- 各景点访问浏览量
- 各景点营业额
- 各景点门票价格



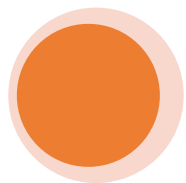
分析结果的可视化

- 技术: Vue框架, ElementPlusUI组件, Echarts组件
- 图表展示: 折线图, 饼图, 直方图

(二) 数据分析——数据流程图



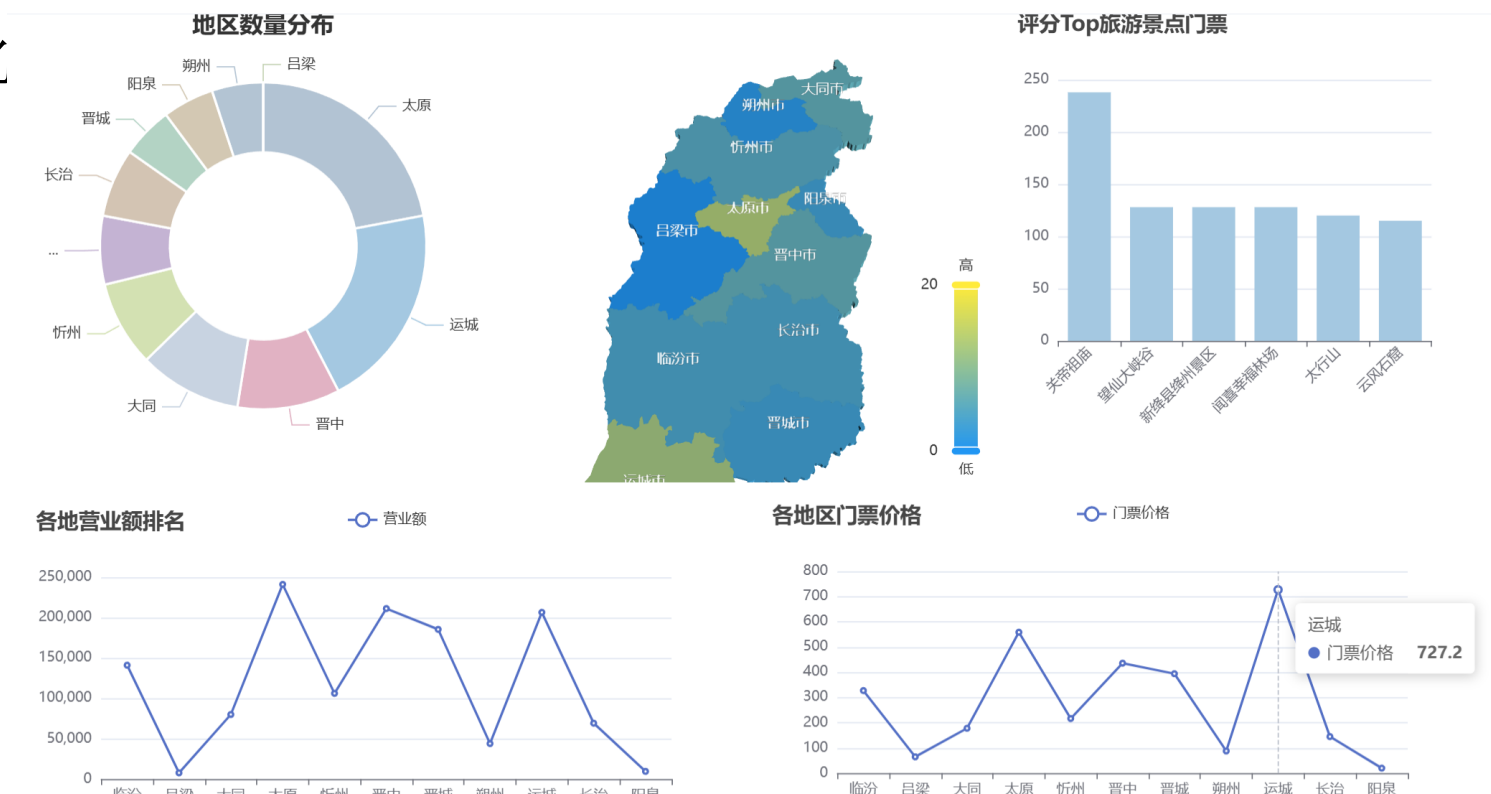
(二) 数据分析——分析结果的可视化



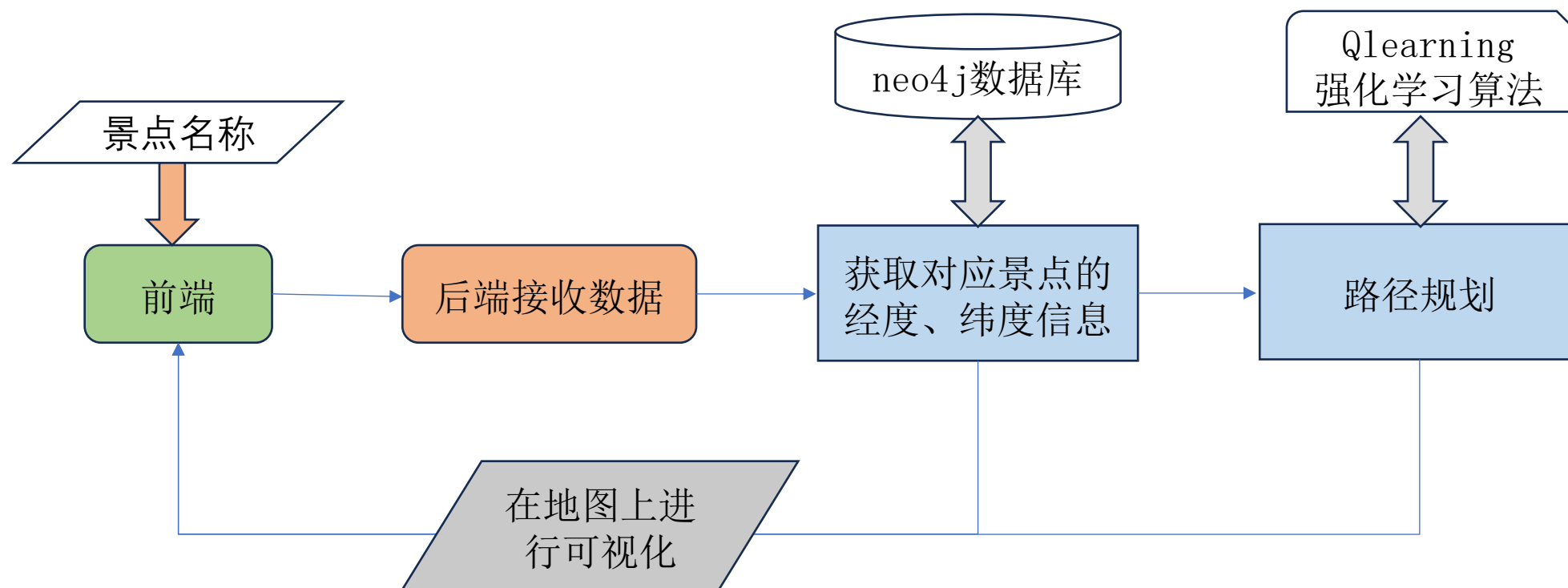
分析结果的可视化

- 技术: Vue框架, ElementPlusUI组件, Echarts组件
- 图表展示: 折线图, 饼图, 直方图

山西景区数据分析可视化



(三) 路线推荐——流程



(三) 路线推荐——输入与反馈

选择旅行策略

请选择旅行策略

选择起点/终点

起点/终点

选择途经点

☐ 五台山

☐ 云冈石窟

☐ 晋祠博物馆

☐ 应县木塔

☐ 山西黄河壶口瀑布旅游区

☐ 平遥古城

☐ ...

搜索路线

旅行策略

起点和终点

途径点



地图展示



具体路线信息

具体路线信息

预计行驶时间 1144 分钟

路线距离 946.0 公里

路径规划

五台山

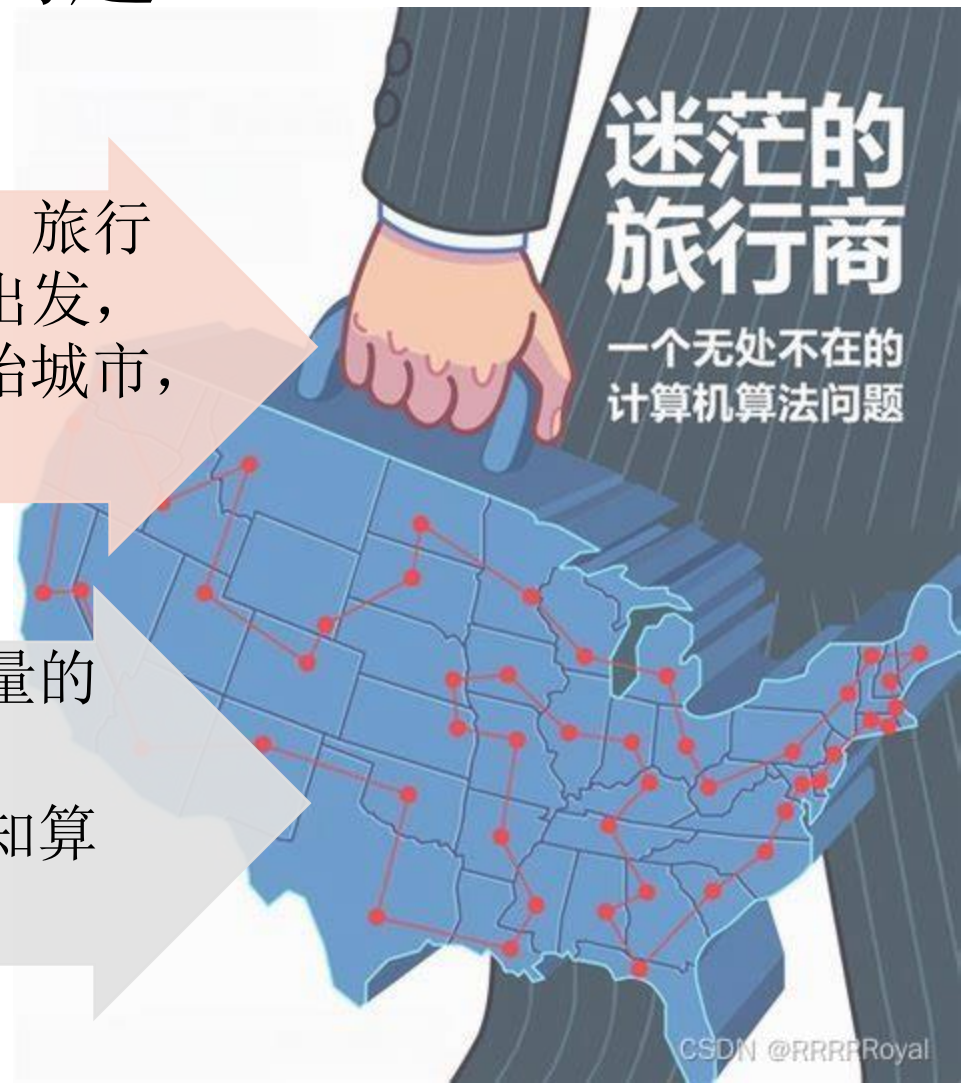
（四）强化学习算法——解决tsp问题

定义

- 给定一组城市和每对城市之间的距离，旅行商的问题问的是旅行商从每一个城市出发，必须访问每个城市一次，最后回到原始城市，这样的最短可能形成是什么？

复杂性

- 组合性：可能的行程数量随着城市数量的增加而阶乘级增加
- NP难性：没有多项式时间复杂度的已知算法可以解决TSP问题



（四）路径推荐算法的探索历程

Dijkstra
最短路径算法

算法思想：基于贪心算法，广度优先遍历核心思想，计算一个节点到其他所有节点的最短路径

GLOP
全局和局部优化策略

算法思想：将大型路由问题划分为旅行商问题 (TSP)，并将 TSP 划分为最短哈密顿路径问题

Qlearning
强化学习算法

算法思想：通过与环境的交互来学习决策策略，以最大化长期累积奖励，智能体（agent）通过执行一系列动作来影响环境，从而获得反馈信号，即奖励（reward）

(四) 强化学习——状态、动作、奖励的定义

状态

- **状态**通常表示旅行商当前所在的城市以及已经访问过的城市集合。
- 位为1表示该城市已被访问，位为0表示该城市尚未被访问。

动作

- 动作在TSP问题中指的是旅行商从当前城市移动到下一个可能城市的决策。
- 动作的选择应基于最小化总旅行距离的目标。因此，动作空间通常包含所有尚未访问过的城市。

奖励

- 奖励在TSP问题中通常定义为负成本，即旅行商从当前城市移动到下一个城市所花费的距离的负值。

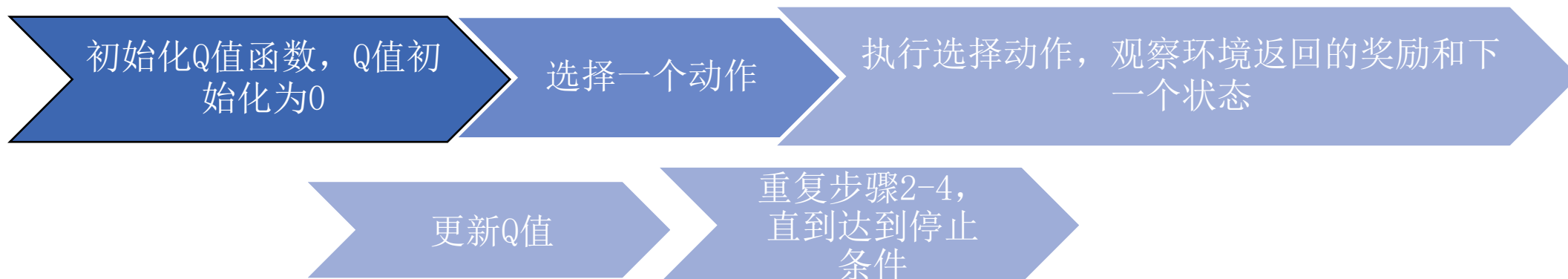
(四) Qlearning强化学习算法策略

简介:

Q-learning是一种强化学习算法，用于解决基于奖励的决策问题。它是一种无模型的[学习方法](#)，通过与环境的交互来学习最优策略。Q-learning的核心思想是通过学习一个Q值函数来指导决策，该函数表示在给定状态下采取某个动作所获得的累积奖励。

训练过程:

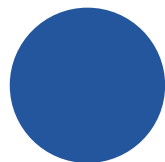
$$R(i, j) = -\alpha \cdot \exp\left(\frac{D(i, j)}{C}\right) + (1 - \alpha) \cdot S(j) - \beta \cdot \log(1 + P(j))$$





项目展示和未来工作

04



前端展示



未来工作展望

前端展示-景区数据分析可视化+路线推荐



太原理工大学
TAIYUAN UNIVERSITY OF TECHNOLOGY

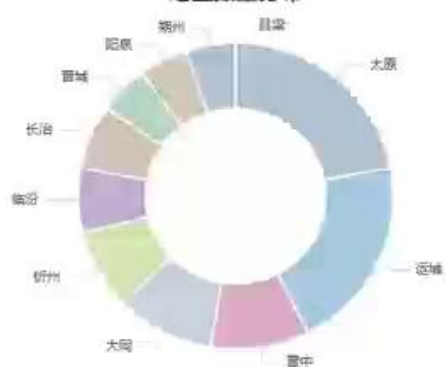


行舟

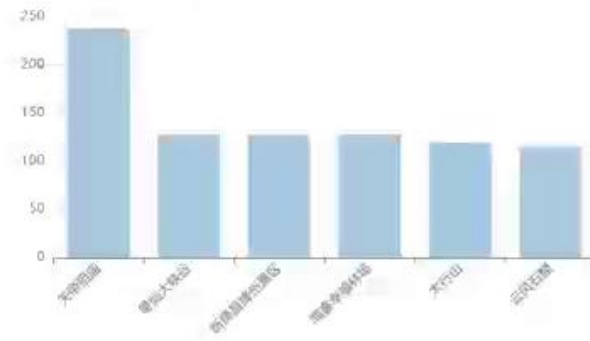
山西景区数据分析可视化

路线规划

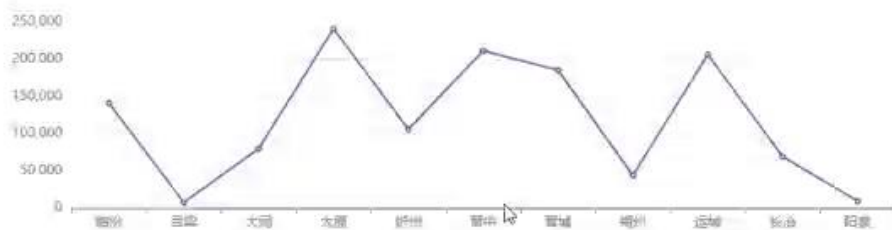
地区数量分布



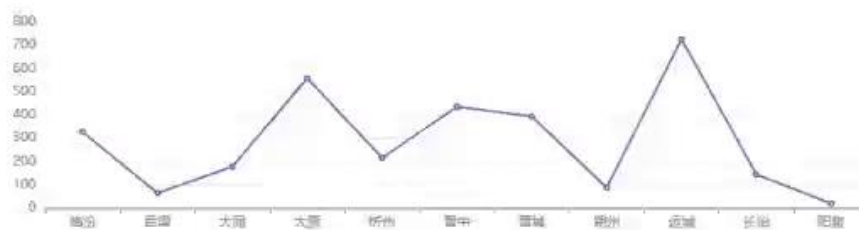
评分Top旅游景区门票



各地营业额排名



各地区门票价格



前端展示-景区数据分析可视化



未来工作展望

数据采集

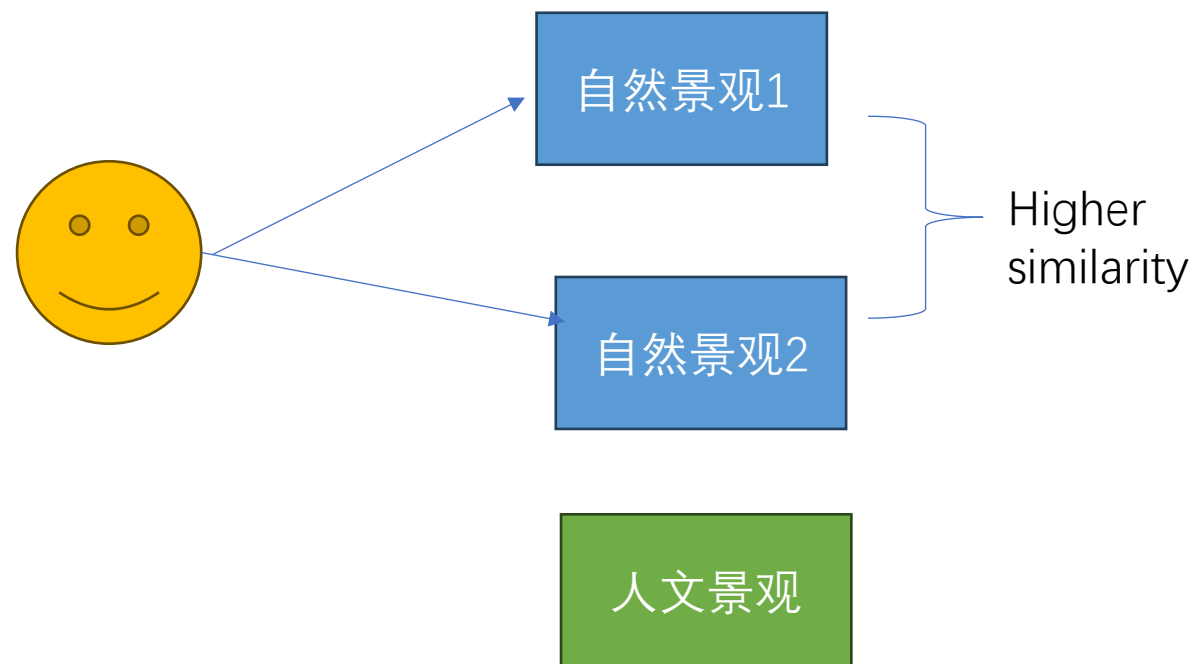
数据源扩展：引入更多数据源来**获取更丰富的用户反馈和兴趣点**。**整合交通数据：**

- 实时路况、交通事故→提供更精确时间预测
- 路段中更多结点信息→细化节点得到更优选择策略
- 路费属性作为奖励函数考虑的一部分→优化行动选择

未来工作展望

个性化推荐算法

开发基于景点划分的推荐模块，根据**景观属性（比如自然/人文）、旅游资源特征等相似度**，提供更精准的推荐结果，后使用路线规划模块,规划路线



未来工作展望

PySpark数据分析深化

大数据处理与分析：利用PySpark的分布式计算能力，开发更复杂的分析模型
时间序列分析、聚类分析、关联规则挖掘。

实时数据分析：研究并实现**实时数据流**处理技术，以支持实时数据分析和决策支持。
开发实时推荐系统，根据**用户的实时行为和兴趣点**，动态调整推荐结果。





太原理工大学
TAIYUAN UNIVERSITY OF TECHNOLOGY



THANKS