



旅游大数据

北京理工大学计算机学院 计卫星

2019年1月



北京理工大学计算机学院
School of Computer Science and Technology, BIT



架构设计与搭建

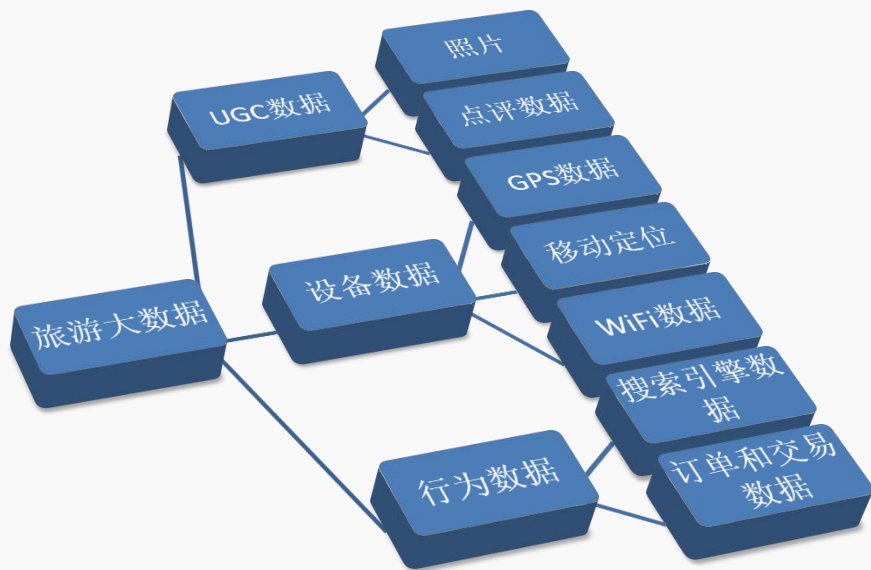


架构设计与搭建

- 基于公众口碑大数据的景区评价
 - 数据抓取、数据整理、数据分析、数据展示
- 架构设计考虑
 - 数据类型：已有评论，将来包括图片、视频等
 - 数据规模：存储总量TB以上
 - 数据分析：分析的工作量比较大
 - 时间要求：不要求实时



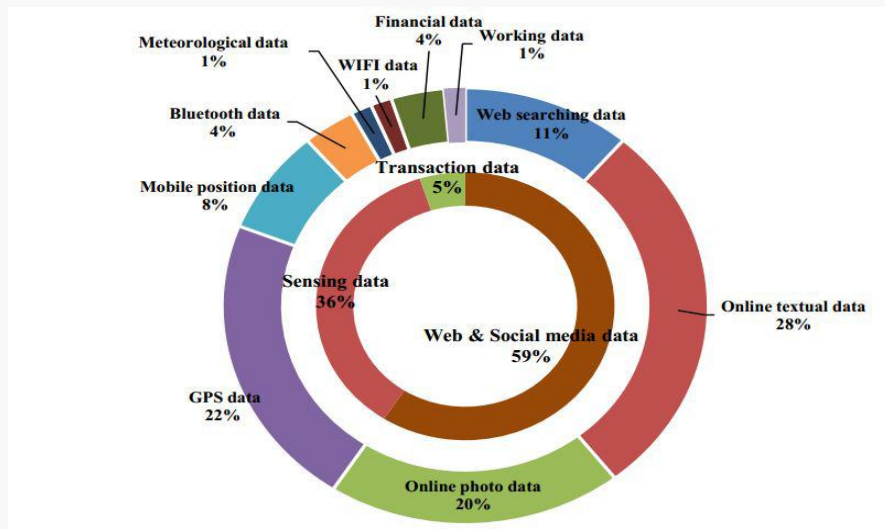
架构设计与搭建





架构设计与搭建

- 数据来源





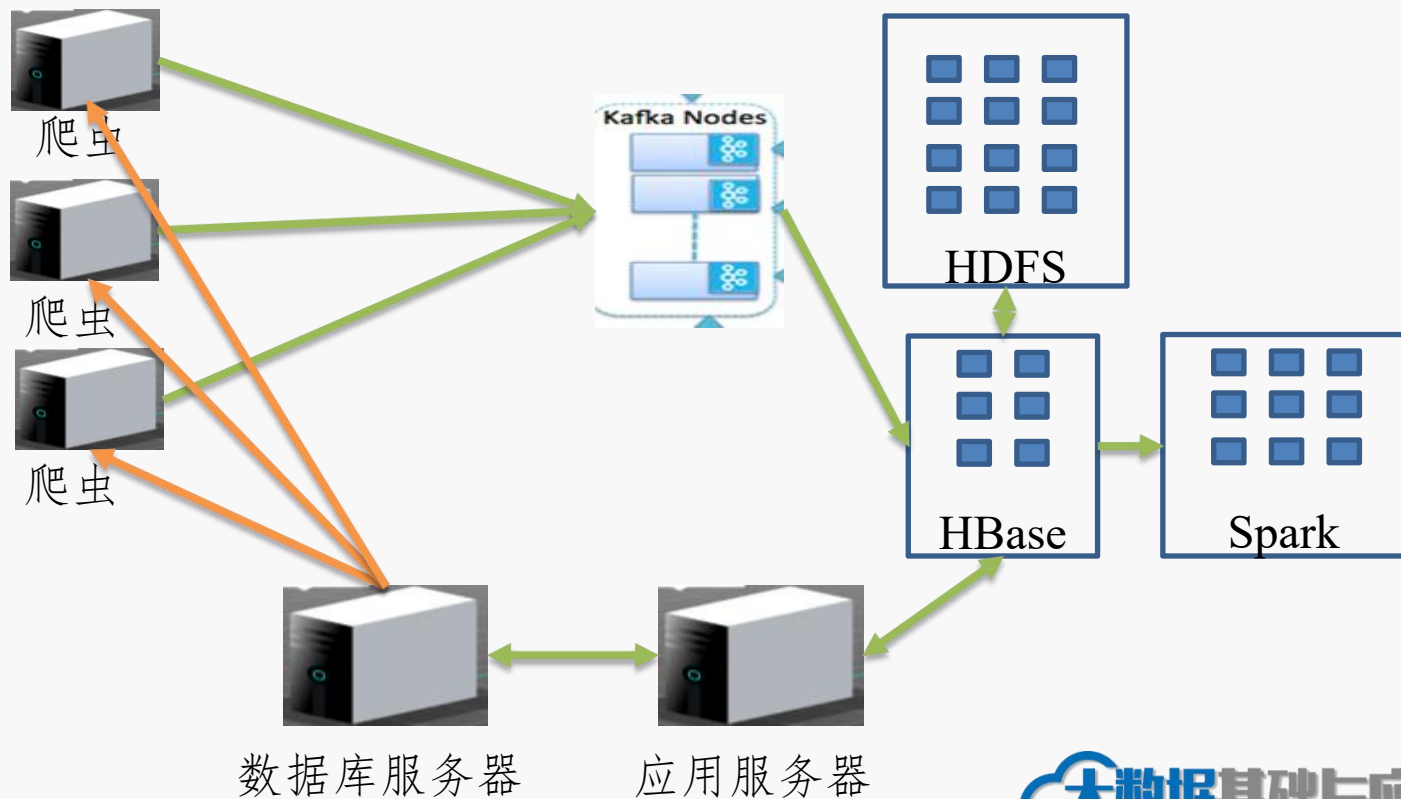
架构设计与搭建

- 基于Hadoop的层次架构设计





架构设计与搭建





数据抓取与存储



数据分析



数据分析

- 评论词云





● 评论词云





数据分析

- 共现矩阵

	白洋淀	荷花	景区	景点	景色	大观园	芦苇荡	门票
白洋淀		198	140	103	71	66	88	40
荷花	198		77	76	76	119	93	43
景区	140	77		65	27	35	30	36
景点	103	76	65			44	36	60
景色	71	76	27					
大观园	66	119	35	44			34	30
芦苇荡	88	93	30	36		34		
门票	40	43	36	60		30		



• 语义网络

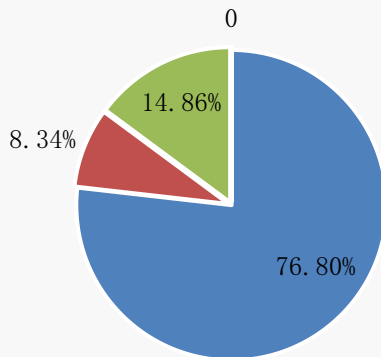




数据分析

- 情感分析

游客情绪占比



■ 积极情绪： ■ 中性情绪： ■ 消极情绪： ■



数据分析

- 景区总体评价

