

数字文旅专业系列教材 · 教学课件

智慧旅游系统架构设计与实践

第5章 中间件与数据平台层（M）——云计算与大数据

授课对象：安顺学院旅游管理学院 · 旅游管理专业本科生

主讲：祝刚

本章导览

- 5.1 数据平台层的架构设计
- 5.2 云计算服务模型与部署
- 5.3 大数据处理技术
- 5.4 数据治理与共享
- 国际比较与经验借鉴
- 技术经济学分析
- 社会文化影响
- 监管框架与政策环境
- 前沿研究方向

- 识记：云计算 IaaS/PaaS/SaaS 服务模型与大数据处理技术栈
- 理解：数据中台理念与数据治理体系
- 应用：设计智慧旅游数据平台架构方案
- 分析：分析不同云部署策略的优劣
- 评价：评估数据平台成熟度与数据质量
- 创造：设计面向省级文旅厅的数据中台方案

- 某省级文旅厅基于阿里云构建旅游大数据平台，汇聚了 OTA 数据、运营商信令数据、景区票务数据和社交媒体数据等 10+ 类数据源。
- 平台采用 MaxCompute 作为离线计算引擎，Flink 作为实时流处理引擎，通过数据中台统一对外提供数据服务。
- Gretzel 等（2015）指出，"智慧旅游明确依赖于不仅收集海量数据，而且智能地存储、处理、组合、分析和使用大数据的能力"[1]。该平台正是这一理念的工程实践。

01

5.1 数据平台层的架构设计

5.1.1 数据平台层在 S.M.A.R.T. 框架中的定位

- 中间件与数据平台层（M）是 S.M.A.R.T. 框架的第二层，承上启下——向下接收感知层（S）采集的多源异构数据，向上为应用层（A）提供数据服务和计算能力。
- Gao 和 Li（2020）在设计基于云服务的智慧旅游管理系统时，提出了四层逻辑架构：模块层、基础层、功能层和展示层 [2]，其中基础层和功能层即对应数据平台层的功能。

5.1.2 数据平台层的功能架构

- 数据平台层的功能架构遵循 " 采集→存储→处理→服务 " 的数据流水线模式：
- 数据采集层。负责从感知层和外部数据源接入数据。石宇良和栗璞（2014）较早提出了基于云计算的智慧旅游架构 [3]，强调了数据采集在整体架构中的基础地位。
- 劳国炜（2021）进一步设计了智慧旅游信息平台数据库建设架构 [4]，包括数据采集系统（网络覆盖与前端硬件选择）和数据处理系统（数据中心架构、数据库选择、门禁票务系统架构）。

5.1.2 数据平台层的功能架构

- 数据存储层。
- 采用混合存储策略——关系数据库（MySQL/PostgreSQL）存储结构化业务数据，NoSQL 数据库（MongoDB/HBase）存储非结构化数据，
- 数据仓库（ClickHouse/Doris）存储分析型数据，图数据库（Neo4j）存储关系型数据。

5.1.2 数据平台层的功能架构

- 数据处理层。包括批处理（Spark/Hive）和流处理（Flink/Kafka Streams）两条处理管线。批处理用于历史数据分析和模型训练，流处理用于实时监控和即时推荐。
- 数据服务层。将处理后的数据以 API、数据可视化、报表等形式对外提供服务。
- Sabbioni 等（2022）提出的服务集成架构 [5] 为数据服务层的设计提供了参考——通过标准化的服务接口支持新型个性化智慧旅游产品。

5.1.3 数据中台理念与实现路径

- 数据中台是近年来在企业级数据架构中兴起的设计理念，其核心是将数据从各业务系统中抽取、清洗、整合，形成统一的数据资产，以服务化的方式提供给上层应用。
- 智慧旅游数据中台的实现路径包括：数据资产盘点（梳理现有数据源和数据流向）、数据标准制定（统一数据命名、格式和质量标准）、数据模型设计（建立旅游领域的统一数据模型，如游客画像模型、景区评价模型、客流预测模型）、数据服务开发（将数据模型封装为 API 供应用调用）和数据治理实施（建立数据质量监控和数据生命周期管理机制）。

02

5.2 云计算服务模型与部署

5.2.1 IaaS 层：虚拟化资源池与弹性计算

- IaaS（基础设施即服务）提供虚拟化的计算、存储和网络资源。智慧旅游系统使用 IaaS 的主要优势是弹性——在旅游旺季自动增加服务器实例应对流量高峰，在淡季自动释放资源降低成本。
- Gao 和 Li（2020）基于云服务的智慧旅游管理系统设计 [2] 验证了 IaaS 在旅游场景中的适用性。

5.2.2 PaaS 层：大数据处理平台与 AI 开发环境

- PaaS（平台即服务）提供应用开发和运行环境。在智慧旅游系统中，PaaS 层承载着大数据处理平台（如 Spark 集群）和 AI 模型训练环境（如 GPU 服务器）。
- Zhang（2022）在研究基于大数据技术的智慧旅游平台构建时 [6]，详细描述了 PaaS 层如何支持旅游大数据的采集、存储、分析和可视化全流程。

5.2.3 SaaS 层：智慧旅游应用服务

- SaaS（软件即服务）以订阅模式提供智慧旅游应用。景区无需自建 IT 团队，按月或按用户数付费即可使用智慧导览、客流分析、智能客服等应用。这种模式特别适合中小型景区。

5.2.4 混合云与多云部署策略

- 大型智慧旅游系统通常采用混合云策略——公有云承载面向公众的应用服务（如游客 App），私有云承载敏感数据（如游客身份信息）。多云部署可以避免供应商锁定，但也增加了运维复杂度。
- 丁俊伟等（2016）在云计算与物联网智慧旅游系统研发中 [7]，采用了混合云架构实现公私数据的分离管理。

03

5.3 大数据处理技术

5.3.1 分布式存储

- 智慧旅游系统产生的数据量大、类型多，传统单机存储无法满足需求。分布式存储技术包括：
- HDFS （ Hadoop Distributed File System ）。适合存储大文件的分布式文件系统，用于存储日志、视频等非结构化数据。
- HBase 。基于 HDFS 的列式 NoSQL 数据库，适合存储传感器时序数据——这类数据写入频繁、读取按时间范围查询。

5.3.1 分布式存储

- MongoDB。文档型 NoSQL 数据库，适合存储 JSON 格式的旅游信息（如景点描述、用户评论）。

5.3.2 批处理与流处理

- Spark。基于内存计算的分布式处理框架，适合迭代式机器学习和交互式数据分析。
- Zhang 和 Cheng（2024）在研究 AI 和大数据如何赋能利益相关者构建智慧旅游目的地时 [8]，使用 Spark 进行旅游大数据的批处理分析。
- Flink。流式处理引擎，适合实时数据处理。在智慧旅游场景中，Flink 用于实时客流监测、异常事件检测和即时推荐。

5.3.2 批处理与流处理

- Bi 和 Liu （2022）的云 IoT 平台 [9] 使用了流处理技术实现旅游信息的实时服务。

5.3.3 数据仓库与 OLAP 分析

- 数据仓库为智慧旅游的决策支持提供分析能力。
- ClickHouse 和 Apache Doris 是两个适合旅游数据分析的 OLAP 引擎——它们支持对亿级数据的秒级聚合查询，可以快速回答 " 过去一个月各景区的日均客流是多少 " 这类分析问题。

5.3.4 数据可视化与 BI 工具

- 数据可视化是将数据分析结果传达给决策者的最后一公里。Du （2021）设计的智能旅游信息系统 [10] 集成了数据挖掘算法和可视化组件，使管理者能够直观地理解旅游数据中的模式和趋势。

04

5.4 数据治理与共享

5.4.1 数据标准与元数据管理

- 数据标准化是数据共享和互操作的基础。劳国炜（2021）在智慧旅游信息平台数据库架构设计中 [4]，强调了数据标准在实现数据便捷高效共享中的关键作用。
- 元数据管理（描述数据的数据）帮助管理者理解数据的来源、格式、质量和使用情况。

5.4.2 数据质量管控体系

- 数据质量管控包括：数据完整性检查（是否有缺失值）、准确性校验（数据是否与真实值一致）、一致性验证（不同系统中的同一数据是否一致）和时效性监控（数据是否及时更新）。
- Rahmadian 等（2023）将大数据治理与可持续旅游关联讨论 [11]，指出数据治理不仅是技术问题，更是治理框架和伦理问题。

5.4.3 数据共享开放机制

- 数据共享是智慧旅游生态系统的核心需求。Yu 和 Wang （2022）在研究智慧旅游生态系统构建时 [12]，提出通过数据共享平台实现政府、企业、游客和社区之间的价值共创。
- 数据共享机制设计需要平衡开放性与安全性——过于开放可能导致数据泄露，过于封闭则限制生态发展。

5.4.4 旅游大数据资产化路径

- 旅游大数据资产化是将数据从 " 成本中心 " 转变为 " 价值中心 " 的过程。
- 资产化路径包括：数据确权（明确数据所有权和使用权）、数据估值（量化数据的商业价值）、数据交易（在合规前提下进行数据买卖）和数据产品化（将数据封装为可销售的数据产品）。

05

国际比较与经验借鉴

云计算在智慧旅游中的全球实践

- 北美智慧旅游平台普遍采用 AWS 或 Azure 作为云基础设施，强调弹性和全球化部署。欧洲由于数据本地化要求（GDPR），更倾向于使用本地云服务商或在欧盟境内部署私有云。
- 中国智慧旅游云平台主要基于阿里云、腾讯云和华为云，石宇良和栗璞（2014）的研究 [3] 是中国较早探讨云计算在旅游信息化中应用的文献。

数据治理的法规差异

- GDPR 要求 " 默认隐私 " 设计，中国的《数据安全法》和《个人信息保护法》要求数据分类分级管理，美国主要依靠行业自律。
- 这些法规差异直接影响数据平台层的架构设计——面向不同市场的系统需要做适应性调整。

06

技术经济学分析

云计算成本优化

- 云计算的按需付费模式虽然灵活，但如果管理不当，成本可能失控。
- 优化策略包括：弹性伸缩（根据负载自动调整实例数量）、预留实例（对长期稳定负载购买预留实例获得折扣）、Spot 实例（对容错性高的批处理任务使用低价 Spot 实例）和存储分层（热数据用高性能存储，冷数据用低成本归档存储）。

数据平台的 ROI 分析

- 数据平台的建设投资通常在百万到千万级别，回报周期 2-4 年。
- ROI 的主要来源：运营效率提升（减少人工数据处理工作）、决策质量改善（数据驱动的决策比经验决策更准确）和新业务机会（数据资产催生新的商业模式）。

07

社会文化影响

数据权力与算法治理

- 数据平台层集中了大量旅游数据，数据平台的运营者实际上掌握了 " 数据权力 "—— 谁有权限访问数据、数据如何被使用、算法如何做出推荐决策，这些问题不仅是技术问题，更是社会治理问题。
- Yu 和 Wang （ 2022 ） 的多中心治理机制 [12] 为数据权力分配提供了理论框架。

数据偏见与公平性

- 旅游大数据中可能隐含系统性偏见——OTA 数据偏向使用在线预订的游客（通常是年轻、高收入群体），可能忽视老年游客和低收入群体的需求。
- 数据平台层的设计应包含偏见检测和缓解机制，确保数据分析结果不会加剧社会不平等。

08

监管框架与政策环境

- 王琼阁（2025）在分析智慧旅游数据安全性与隐私保护时指出 [13]，" 大数据技术的广泛应用带来的数据汇聚虽为旅游服务的优化提供了机遇，但也带来了安全上的挑战 "。
- 数据平台层的合规设计需要满足《网络安全法》的等保要求、《数据安全法》的数据分类分级要求，以及《个人信息保护法》的最小必要原则。

- 面向国际游客的智慧旅游系统可能涉及数据跨境传输。《数据安全法》要求重要数据出境必须通过安全评估，这要求数据平台层在设计时就要考虑数据本地化存储和跨境传输的合规路径。

09

前沿研究方向

数据网格（Data Mesh）

- 数据网格是一种去中心化的数据架构理念，将数据所有权和管理权下放到各业务域，而非集中到中央数据团队。
- 在智慧旅游场景中，票务数据由票务团队管理，客流数据由运营团队管理，每个团队以 " 数据产品 " 的形式对外提供数据服务。这种模式减少了对中央数据团队的依赖，提高了数据响应速度。

隐私计算

- 隐私计算技术（联邦学习、多方安全计算、同态加密）使数据可以 " 可用不可见 "—— 各方可以在不共享原始数据的情况下进行联合分析。
- 这对于旅游行业的数据共享具有重要价值——不同企业可以在不暴露商业敏感数据的前提下进行联合营销分析。

- 中间件与数据平台层（M）是 S.M.A.R.T. 框架的数据中枢，负责数据的存储、处理、治理和服务化。
- 本章阐述了云计算 IaaS/PaaS/SaaS 服务模型在智慧旅游中的应用，介绍了分布式存储、批处理与流处理、数据仓库与 OLAP 分析等大数据技术，讨论了数据标准、
- 数据质量管控、数据共享和大数据资产化等治理议题。

- （识记） 数据平台层的四个功能子层是什么？
- （理解） 比较 IaaS、PaaS 和 SaaS 在智慧旅游中的适用场景。
- （应用） 为一个拥有 10 个景区的市级文旅集团设计数据中台架构方案。
- （分析） 分析批处理（Spark）与流处理（Flink）在智慧旅游中的互补关系。
- （评价） 评估数据网格（Data Mesh）理念在智慧旅游领域的适用性。
- （创造） 设计一个基于隐私计算的旅游企业间数据共享方案。

- [1] Gretzel U, et al. Smart tourism: foundations and developments[J]. Electronic
- [2] Gao F, Li L. Design of Smart Travel Management System Based on Cloud Service
- [3] 石宇良, 栗璞. 基于云计算平台的旅游信息化的研究 [J]. 维普, 2014.
- [4] 劳国炜. 智慧旅游信息平台数据库建设和架构设计 [J]. 维普, 2021.
- [5] Sabbioni A, et al. An Architecture for Service Integration to Fully Support
- [6] Zhang Y. Research on the Construction and Application of Smart Tourism Platf

谢谢聆听 · 下章再见

安顺学院旅游管理学院 · 数字文旅专业系列教材