

基于SQL Server的在线教育数据仓库系统的设计与实现

朱怡菁, 曹红苹*, 吴亦晟, 陈 功

上海工程技术大学管理学院, 上海

收稿日期: 2025年6月14日; 录用日期: 2025年7月12日; 发布日期: 2025年7月18日

摘 要

本研究基于SQL Server构建在线教育数据仓库系统, 采用ETL技术实现多源数据集成, 运用星型/雪花混合模型组织数据, 结合SSAS和SSRS实现多维分析与可视化。该方案有效整合分散的教育数据, 通过智能分析优化课程设计, 为管理者提供决策支持。

关键词

数据仓库系统, SQL Server, 在线教育, 多维数据分析

Design and Implementation of Online Educational Data Warehouse System Based on SQL Server

Yijing Zhu, Hongping Cao*, Yisheng Wu, Gong Chen

School of Management, Shanghai University of Engineering Science, Shanghai

Received: Jun. 14th, 2025; accepted: Jul. 12th, 2025; published: Jul. 18th, 2025

Abstract

This research builds an online education data warehouse system based on SQL Server, adopts ETL technology to achieve multi-source data integration, uses the star/snowflake hybrid model to organize data, and combines SSAS and SSRS to realize multi-dimensional analysis and visualization. This scheme effectively integrates scattered educational data, optimizes curriculum design through

*通讯作者。

文章引用: 朱怡菁, 曹红苹, 吴亦晟, 陈功. 基于SQL Server的在线教育数据仓库系统的设计与实现[J]. 计算机科学与应用, 2025, 15(7): 56-61. DOI: 10.12677/csa.2025.157180

intelligent analysis, and provides decision support for managers.

Keywords

Data Warehouse System, SQL Server, Online Education, Multi-Dimensional Data Analysis

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



1. 引言

在线教育行业随着互联网技术的飞速突破而发展迅速，用户规模和市场规模不断扩大。然而伴随着数据量的增加，传统数据库系统在处理大量复杂数据时面临性能瓶颈和数据集成困难[1]。因此，本研究的目的是开发并实现一个基于 SQL Server 的在线教育数据仓库系统，通过 ETL 技术实现多源数据集成，采用星型/雪花模型组织数据，并利用 SSAS 进行挖掘分析、SSRS 生成可视化报表。

现有研究如刘力[2]采用纯星型模型简化查询，但牺牲了维度分析的灵活性，而基于 Hadoop 的方案虽支持海量数据，但实时性不足。本研究的创新性在于：混合模型优化，结合星型模型与雪花模型的优势，在查询效率与深度分析之间取得平衡；技术栈整合，首次在 SQL Server 生态中实现 SSIS、SSAS、SSRS 全链路闭环；教育场景适配，针对学习行为分析设计专用度量。

2. 系统设计

2.1. 系统总体架构设计

本系统总体架构采用分层设计理念，自下而上依次为数据采集层、ETL 层、数据存储层、数据分析层和报表展示层，各层紧密协作，共同实现在线教育数据的高效管理与分析[3]。

2.2. 系统功能设计

数据采集模块支持多源异构数据接入，包括 SQL Server 关系型数据库和 CSV/Excel 文件系统，实现学生学习行为、课程资源等数据的自动化采集[4]。通过 SSIS 工具建立数据管道，确保数据抽取的高效性和准确性[5]。

数据存储模块采用星型 - 雪花混合模型组织数据，构建以学习行为和教学效果为核心的事实表，关联维度表[6]。通过建立聚集索引和分区存储优化查询性能[7]。

数据分析模块基于 SSAS 实现多维 OLAP 分析，支持切片、切块、钻取等分析行为。

报表生成模块通过 SSRS 提供可视化设计界面，支持拖拽式创建各类型报表。

2.3. 数据仓库设计

2.3.1. 数据仓库概念设计

本系统采用多维数据模型架构，围绕“学生学习行为分析”和“教学效果评估”两大主题设计[8]。事实表包括学习行为事实表(记录学习时长、视频观看时长等指标)和考试结果事实表(成绩、答题时间等指标) [9]。维度表设计涵盖时间维度表(支持多级时间分析)、课程维度表(实现课程层次钻取)和学生维度表(用户细分分析)，通过星型 - 雪花混合模型实现高效查询与深度分析。

2.3.2. 数据仓库逻辑设计

在线教育数据仓库采用星型 - 雪花混合模型, 在保证查询效率的同时支持深度分析[10]。维度表设计方面, 课程维度表(DimCourse)以代理键 CourseKey 为主键, 包含课程 ID、标题、状态和创建时间等核心属性。用户维度表(DimUser)通过 UserKey 唯一标识用户, 记录用户 ID、角色和等级信息。时间维度表(DimTime)提供完整的时间层次结构, 包含日期、季度等字段。章节维度表通过 ChapterKey 关联课程维度, 记录章节编号、标题等详细信息。任务维度表则与章节维度关联, 包含任务编号、标题、时间范围等属性。试卷维度表(DimTest)以 TestKey 为主键, 关联课程维度, 记录试卷名称、限时、总分等考试特征信息[11]。

学习行为事实表(FactLearning)通过 UserKey、CourseKey、TaskKey 和 TimeKey 关联各维度表, 记录学习时长、视频观看时长和任务完成状态等核心指标。考试结果事实表(FactTestResults)同样采用多维度关联设计, 存储考试成绩、答题时间、通过状态和教师评语等关键数据。

2.3.3. 数据仓库物理设计

该数据仓库采用星型 - 雪花混合存储架构, 建立“主键聚集 - 外键覆盖 - 热点组合”三级索引体系。基于访问频率实施智能分层存储: 热数据(3 月内)使用 NVMe SSD, 温数据存 SAS HDD, 冷数据归档至磁带库。通过时间分区函数实现季度数据划分, 优化存储成本与查询性能的平衡。

3. 系统实现与应用

本研究基于天池平台提供的在线教育数据集, 链接为 <https://tianchi.aliyun.com/dataset/172411>, 采用分层处理策略实现数据采集。首先通过 Excel 进行编码和格式转换, 再利用 Power Query 完成数据清洗预处理, 最终通过 SQL Server 实现高效存储。对于大规模数据, 采用 SSIS 包实现自动化批量导入, 构建完整的数据处理链路。

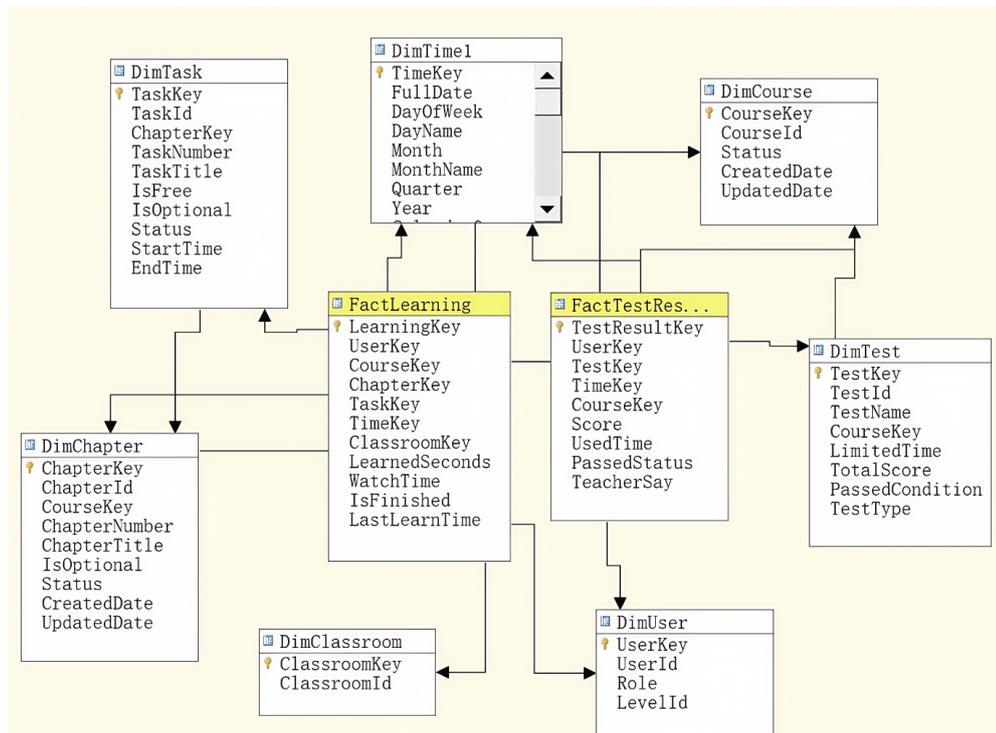


Figure 1. Multi-dimensional dataset

图 1. 多维数据集

数据分析模块在 SSDT 中实现完整的 ETL 流程。通过创建 SSIS 项目, 使用 OLE DB 组件连接数据库, 经数据转换和条件拆分等处理后加载至数据仓库。同时, 构建 Analysis Services 项目, 创建维度和多维数据集, 配置层次结构与度量值, 完成 OLAP 立方体建模, 见图 1。该系统支持上卷、下钻、切片和钻取四种核心分析操作, 为在线教育提供全方位的分析视角。上卷操作通过减少维度或爬升概念层次结构来合并或聚合数据, 从而从更宏观的角度观察学习情况。下钻操作通过将信息分解为更小的部分来深入分析数据。例如深入到按月份、按具体日期的数据进行查看, 帮助用户发现更细粒度的数据细节。切片操作允许用户按时间、课程、学生等多个维度进行灵活操作, 揭示不同因素间的关联关系。钻取操作支持用户从宏观数据深入到微观细节, 例如从课程的总体学习指标逐步钻取到每个学生的具体学习记录, 包括学习时长、完成的任务、考试成绩等详细信息, 见图 2。针对在线教育的特点和需求, 系统支持聚类分析、关联规则挖掘等数据挖掘算法。在聚类分析中, 通过设置合适的参数, 如聚类的数量、距离度量方式等, 将学生按照学习行为模式、成绩水平等因素进行分类, 发现不同类型学生的特点和共性, 为个性化教学和课程推荐提供依据。在关联规则挖掘方面, 通过设置支持度、置信度等参数, 输入相关的数据, 运行数据挖掘模型, 挖掘学生学习行为与课程成绩之间的潜在关系。

平均值项:Score	列标签	2020						
班级		1	2	3	4	5	6	7
1		77.00588235						
2		71.70714286	76.7					
3			73.59					
4		74.56363636		84.425				
5			72.55925926					
6				74.85666667				
7					73.92272727			
8					70.63333333	72.43181818		
9						76.0375	74.80769231	
10							80.90555556	

Figure 2. Results of drilling operations

图 2. 钻取操作结果

数据展示模块通过 Reporting Services 项目实现可视化呈现。连接 SSAS 多维数据集作为数据源, 使用 MDX 查询构建共享数据集。在报表设计中采用表格控件展示学习详情, 可以按课程和时间维度展示学生学习时长分布, 实现多维度数据的有效呈现, 能够动态计算各维度组合的汇总值与占比, 直观揭示学习行为的时间规律与课程偏好, 帮助用户从宏观角度把握学生学习情况的分布特征。例如, 通过分析“学习行为事实表”中的时间维度数据, 发现学生在 10:00~12:00 的活跃度最高, 可建议将重要课程或直播教学安排在该时段。同时, 通过对比不同课程的学习时长和完成率, 筛选出学生普遍投入时间较少的课程, 为优化课程内容或调整教学节奏提供依据。图表报表可以通过可视化的方式展示学生学习行为趋势, 如折线图、柱状图等。将字段分别拖到类别区域和值区域, 并设置图表样式, 包括颜色、图例、标题等, 以直观的图表展示数据趋势, 如按月统计课程参与人数的波动情况, 或对比不同地区学生的平均成绩差异, 强化数据的时空洞察力, 使用户能够更直观地理解数据背后的含义, 见图 3。例如, 通过柱状图发现某班学生的平均成绩显著低于其他班级, 可针对性地加强该班级的教学辅导; 或通过折线图分析某课程的学习时长随时间下降的趋势, 及时调整教学策略。报表通过字段绑定和样式优化, 支持数据下钻查看明细。例如从整体课程表现钻取到具体章节的学习数据, 帮助教师识别难点内容并调整教学重点。最终将设计完成的报表发布至 SSRS 服务器或 SharePoint 平台, 实现在线共享与访问, 为教育决策提供有力支持。

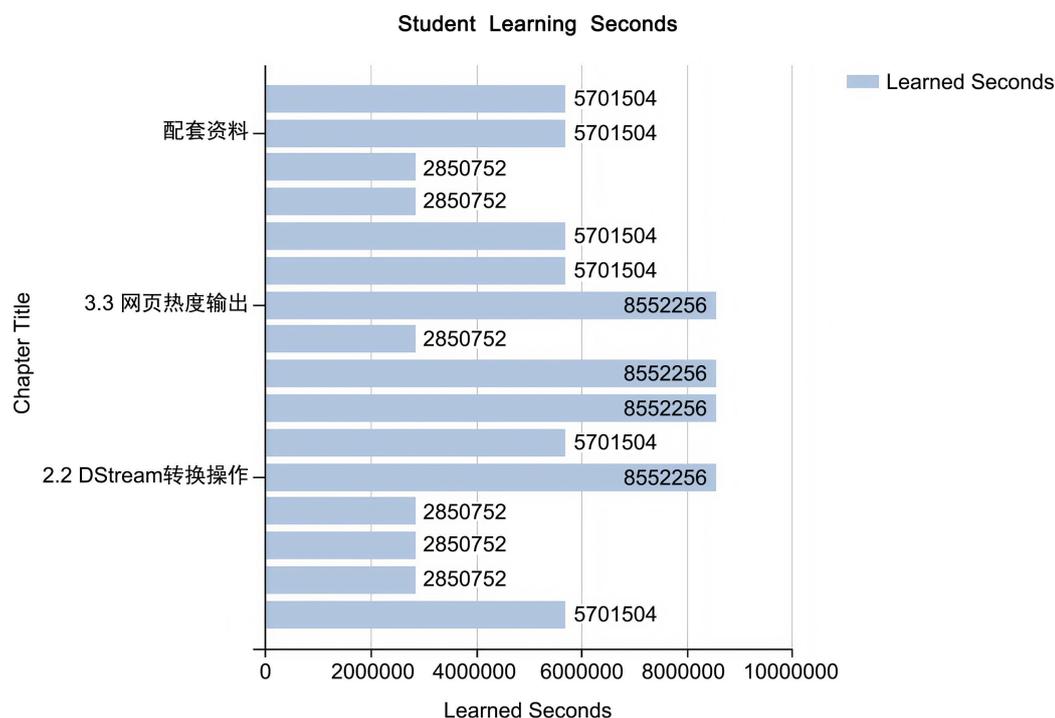


Figure 3. Statistics of chapter learning duration

图 3. 章节学习时长统计

系统测试验证了三大核心模块功能：ETL 模块测试确认数据抽取、清洗和加载流程运行正常，数据处理准确无误；多维数据建模测试验证了维度表与事实表的关联关系正确，MDX 查询响应快速且结果精确；数据展示模块测试确保 OLAP 分析操作(切片、钻取等)功能完整，SSRS 报表数据呈现准确、布局合理。所有测试结果均符合预期，系统各模块运行正常。

4. 总结与展望

本研究基于 SQL Server 技术栈构建了在线教育数据仓库系统，通过 ETL 流程、星型 - 雪花混合模型及多维分析技术，实现了教育数据的有效管理与智能分析。系统整合了 SSIS 数据集成、SSAS 多维建模和 SSRS 可视化报表，为教学决策提供了全面支持。

未来研究可从三个方向拓展：引入大数据技术处理海量教育数据；开发实时分析模块；结合机器学习算法，实现个性化学习推荐。这些改进将进一步提升系统的数据处理能力和智能化水平。

参考文献

- [1] 周舒. 基于商务智能的数据分析平台设计与实现[D]: [硕士学位论文]. 济南: 山东大学, 2019.
- [2] 刘力. 在线教育行业数据仓库系统的设计与实现[D]: [硕士学位论文]. 武汉: 华中科技大学, 2024.
- [3] 刘聪. 基于 Hadoop 的在线教育分析平台的设计与实现[D]: [硕士学位论文]. 上海: 华东师范大学, 2023.
- [4] 周鹏. 基于大数据平台的 K12 在线教育数据仓库设计与实现[D]: [硕士学位论文]. 廊坊: 北华航天工业学院, 2020.
- [5] 王焕民. 基于教育大数据的在线学习分析系统构建研究[J]. 中国现代教育装备, 2019(15): 14-16, 20.
- [6] 白宇斐. 某企业数据仓库的设计与实现[D]: [硕士学位论文]. 北京: 北京邮电大学, 2021.
- [7] 李鹭成, 潘建江. 基于 Power BI 的电商企业数据仓库可视化平台设计与应用[J]. 无线互联科技, 2025, 22(3): 98-101.

-
- [8] 郭云鹏, 吴娟, 陆映梅, 等. 基于数据仓库的电网全业务系统审计数据分析系统设计[J]. 数字技术与应用, 2024, 42(6): 60-62.
- [9] 刘刚. 基于数据仓库的网络成人教育教学系统设计[J]. 信息与电脑(理论版), 2024, 36(5): 245-247.
- [10] Zhou, J. and Li, X. (2024) Application of Big Data in College Student Education Management Based on Data Warehouse Technology and Integrated Learning. *International Journal of e-Collaboration (IJeC)*, **20**, 1-20.
<https://doi.org/10.4018/ijec.346368>
- [11] Fabijanić, M., Ružak, D., Novosel, A. and Hlupić, T. (2023) Data Warehouse-Based Analytical System in Private Higher Education Institution. 2023 46th MIPRO ICT and Electronics Convention (MIPRO), Opatija, 22-26 May 2023, 1287-1292.
<https://doi.org/10.23919/MIPRO57284.2023.10159970>