

Connecting LMS with Data Platform to Break Log Retention Limits and Enable Cross-Year Learning Analytics

Kankan Chen

Learning Mall

Supported by LM

1. Background

The university's online learning platform, **Learning Mall Core**, is built on Moodle and provides course learning, assignments, quizzes, and teaching interactions for all staff and students. As the platform operates on an "academic year cycle," the system from each year is archived at the end of the term to provide a relatively "clean" environment for the new academic year to host new courses and activities.

With the platform's increasing scale of use, the number of courses, access volume, and log data have grown rapidly. To ensure production system performance and storage space, log tables (such as learning behavior log tables) in Learning Mall Core were originally set to retain data for only the last **180 days**. Consequently, once the 180-day window passed, teachers, schools, and functional departments struggled to

obtain complete historical behavioral data within the LMS itself, let alone conduct **cross-academic-year and cross-cohort learning analytics and teaching quality assessments** within a unified data view.

Simultaneously, the university's IT department established a unified data middle platform to aggregate data assets from various business systems such as academic affairs, teaching, and research. The challenge became how to integrate Learning Mall's key learning data into this **data middle platform** without disrupting the existing annual archiving mechanism and log retention policy, thereby addressing the pain points of "logs retained for only 180 days" and "system archiving causing data fragmentation." This became a joint problem for the Learning Mall department and the IT data team to solve.

2. Solutions

Adhering to the principle of "maintaining business rhythms while upgrading data capabilities," this case accomplished solution design and implementation through the following aspects:

1. Defining the Learning Analytics Data Model and Scope

- The Learning Mall department and IT data team jointly identified data tables and fields within the LMS crucial for learning analytics, focusing on:

- User and organizational information (user tables, school/major/class associations, etc.).
 - Course and course categorization (course tables, semester information, etc.).
 - Course enrollment and role data (teacher/student roles, enrollment relationships).
 - Grades and assessment data (quiz scores, assignment scores, etc.).
 - Behavior log tables (access, clicks, activity participation records, etc.).
- Based on this, a Learning Mall learning data domain model was established, providing a consistent data foundation for subsequent data middle platform modeling and synchronization.

2. Completing Batch Integration with the Data Middle Platform in Spring 2024

- Defined the initial scope of data to be integrated, focusing on user, course, enrollment, and log data.
- Performed a one-time historical data backfill import, migrating key learning data from previous academic years into the data middle platform.

- Subsequently, implemented scheduled incremental synchronization processes (daily or batch) to ensure continuous updates of Learning Mall data within the data middle platform.

3. Designing a Layered Storage Mechanism: "180-Day Logs + Long-Term Middle Platform Retention"

- The production environment continues its "180-day log retention" policy to ensure Moodle system stability and database performance.
- A dedicated Learning Mall data subject was created within the data middle platform, utilizing high-capacity storage and cold/hot data tiering strategies to enable long-term preservation and cross-year querying of learning behavior logs.
- This created a clear architectural division: "LMS responsible for teaching operations, Data Middle Platform responsible for long-term data retention and analytical computation."

4. Establishing Standardized Data Service and Access Mechanisms

- Created unified data service interfaces (e.g., views, APIs, datasets) for Learning Mall data within the data middle platform for use in future learning dashboards, statistical reports, and specialized analyses.

- Aligned with data governance requirements, configured different data access permissions based on user roles, preventing direct access to the LMS production database.

5. Real-World Application for Teaching: The MTH007 Course Case

- In October 2024, the instructor for the MTH007 course requested online learning behavior data for students from the S1 semester of the 2023/24 academic year for course summarization and teaching improvement.
- Under the original "180-day log retention" policy, the Moodle production system could no longer directly provide complete behavior logs for that semester.
- Leveraging the LM-Data Middle Platform integration completed in Spring 2024, the LM team and IT data team successfully extracted and organized MTH007 learning behavior data for 23/24 S1 (including login counts, resource access, quiz participation, etc.) via the data middle platform. This provided the instructor with cross-academic-year learning analytics support, serving as a typical implementation case of this solution in a frontline teaching scenario.

3. Outcomes and Benefits

- **Broke Through Log Retention and System Archiving Limits, Enabling Long-Term, Continuous Data Accumulation:** Even as the LMS production environment maintains its "180-day log retention, annual archiving" operational policy, Learning Mall learning data within the middle platform can be continuously accumulated across years and cohorts. Teaching management departments and schools can conduct longitudinal comparative analyses over multiple academic years within a unified data view, supporting course development evaluation and teaching reform effectiveness assessment.
- **Validated by Real Teaching Needs: Supported Cross-Year Analysis for MTH007:** In the actual case of MTH007, relying solely on the original LMS logs could not meet the instructor's need to query behavior data from the previous academic year's S1 semester. Using the data middle platform, learning behavior logs for 23/24 S1 were successfully extracted and analyzed, enabling quantitative presentation of student learning activity, resource usage, and quiz participation, providing the instructor with evidence-based material for teaching reflection. This case not only solved a specific course's problem but also allowed the instructor to tangibly experience the direct support and value of data middle platform development for teaching.

- **Significantly Enhanced Decision Support and Data Application**

Capabilities: University leadership, schools, and functional departments can construct multi-dimensional learning dashboards and monitoring indicators based on the data subjects in the middle platform, gaining timely insights into online learning situations and trends across schools and courses. This provides the necessary historical data foundation for future development of student learning early warning models and teaching quality diagnosis models.

- **Reduced LMS Operational Pressure, Optimized Overall System**

Architecture: The Moodle production database no longer bears the burden of long-term log storage and complex historical queries, allowing it to focus more on stably supporting teaching operations. The data middle platform uniformly handles the historical data storage and analytical computation tasks for Learning Mall Core, clarifying architectural responsibilities and facilitating future integration of data from more teaching systems.

- **Improved Data Governance, Security, and Compliance:** Centralized management of access permissions and data masking strategies through the data middle platform avoids the fragmented practice of "each system pulling data directly from the database." This provides a practical example for improving the university's overall data governance framework.

4. Replicability and Promotion Value

1. Applicable to Similar LMS + Academic Year Archiving

Scenarios: For other universities also using Moodle or other LMSs with annual archiving and log retention time limits, this case offers a general pathway of "upgrading data capabilities via a data middle platform without changing teaching business rhythms."

2. Applicable to Data Middle Platform Integration for Other Teaching

Business Systems: This case has validated the complete process from mapping a business system's data model → defining integration scope → establishing a middle platform data domain → designing layered storage and data services. This model can be extended to other teaching business systems like academic affairs systems or continuing education platforms, gradually forming a "teaching big data foundation" centered on the data middle platform.

3. Laying the Groundwork for Future AI Learning Analytics and

Intelligent Applications: Long-term, continuous learning behavior data is a prerequisite for developing AI-powered learning analytics, personalized recommendations, risk warnings, etc. The implementation of this case provides a solid data foundation for future exploration in areas like learning analytics agents and

intelligent dashboards, demonstrating significant potential for extension and promotion.

5. Next Steps

1. Developing Multi-Role Learning Analytics Visualization

Dashboards: Based on the accumulated Learning Mall Core data in the data middle platform, further design and develop learning analytics visualization dashboards tailored for different roles (university leadership, school deans, teachers, etc.), providing tiered, drill-down online learning analysis views.

2. Exploring Student Learning Early Warning and Teaching Quality

Assessment Models: While ensuring data security and privacy, explore building student learning risk warning models and course quality assessment models based on historical learning behavior data, providing more proactive decision support for academic affairs and schools.

3. Deep Integration with Intelligent Learning Analytics

Agents: Aligning with the university's planned intelligent learning analytics agent initiative, utilize the Learning Mall Core learning data in the data middle platform as the underlying data source. This will provide teachers and administrators with "Q&A-style" instant learning insights, further unlocking the value of the data.