A Data Lake for Educational Data Mining

Improving Faculty's Teaching, and Students' Learning

Stephan Ao, Ph.D.

In the past, institutions used traditional data pipelines for analysis: collect, move, clean, and analyze. Today, as massive data is produced on the daily basis, such analytic methodology is labor-intensive, resource-extensive, and time-consuming.

Big Data concepts have been tossed around in recent years in every industry. But not many people truly appreciate the power of Big Data technology, as they stop at using the nomenclatures only, not diving into how the technology works.

Big Data doesn't mean voluminous datasets. In Big Data technology, we don't move, clean, nor change the raw data we collect and store. Better yet, real-time data can be live-streamed and analyzed as events occur. This would not be possible without the newly developed big data frameworks, high-speed communication networks, and cost-effective computing power.

Figure 1 illustrates a big-data-based solution architect for EDM and timely analytics to faster realize an institution's vision through layered mission implementation.

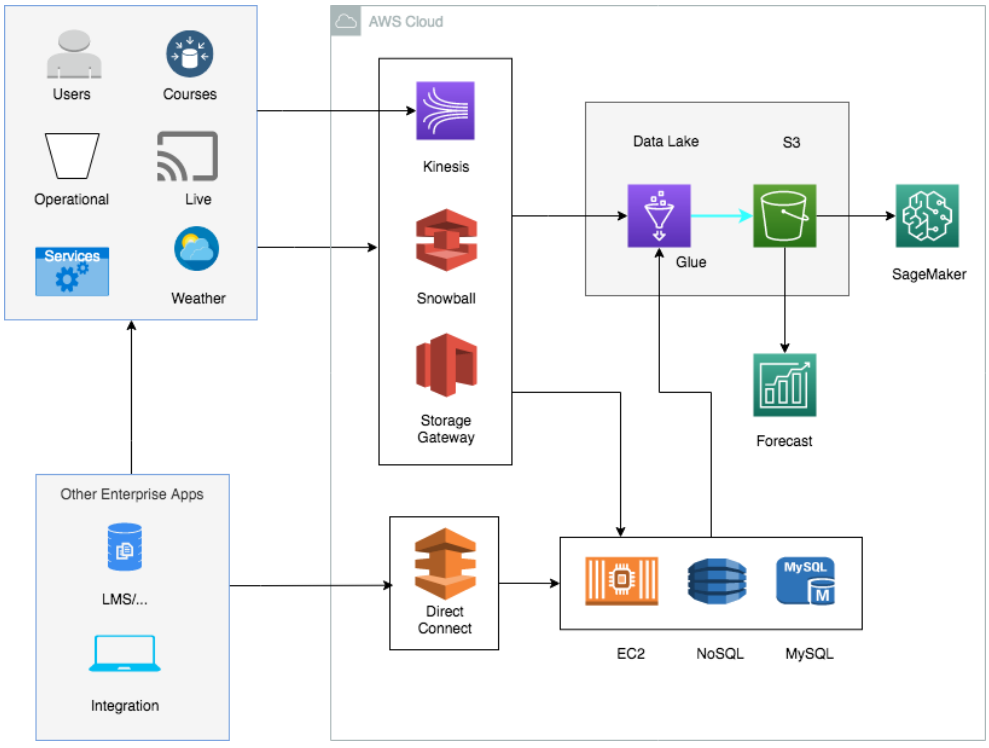

Fig. 1 Educational Data Mining analytics implemented on AWS

For the higher education industry, applications of Big Data technology can speed up innovation in redesigning new or improving existing educational systems. Specifically, for a higher education institution, what can be done to fulfill her mission: to redesign or improve its educational system and educate a new generation of talents for the changing society?

The job is twofold: to collect data in a sensible way, and to put in use of the collected data in daily operations at institution & school levels as each school has her own mission. The following list of data sources is not exhaustive. Not every data source will be utilized in an educational data mining (EDM) application.

User-related data sources:
- Students & profiles
- Teaching faculty & profiles, research projects
- Administration teams
- Management teams
- External contractors & Consultants
- Alumni & External users

Course-related data sources:
- Courses offline & online
- Teaching environment and teaching aids
- Assignments, quizzes, tests, and exams
- Evaluations
- Mobile and web Apps and usage logs

Operational data:
- Management, policies, and rules
- LMS and logs
- Finance and financial aids
- Transactions
- Certifications and diplomas
- Events
- Calendar and holiday logs

Live streaming data:
- Cameras in the classroom and on campus
- Other sensors and IoTs
- Wearable devices
- Others

Facilities & services
- Library collections & services
- Sports ground and activities
- Food courts & services
- Medical services
- Local business & services

Following analytic applications can be developed by interested faculties indifferent Schools:

1. Further innovate the current educational system according to desired institution KPIs. Figure 2 shows a close loop continuous process for improvement.

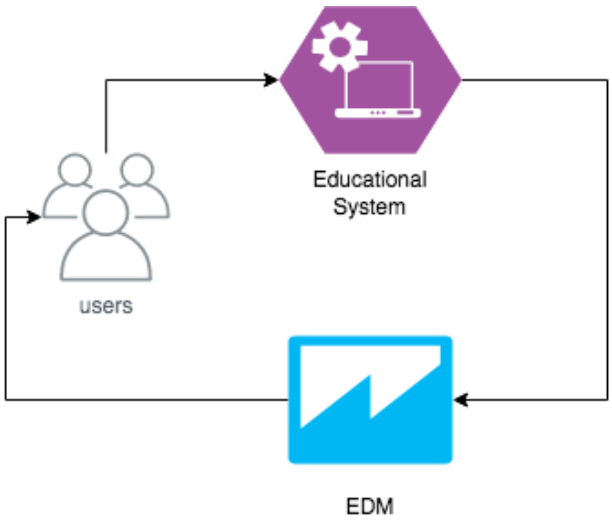

Fig.2. EDM analytics improves teaching and learning

Users: all institute users
Educational System: on-campus and virtual classrooms, LMSs, T& L activities
EDM: implementation of the solution architect and analytic models

2. Student Modeling:
   Student state (frustrated, motivated, satisfied)
   Learning style (VARK)
   Learning progress
   Predict success or failure
   Early intervention for success

3. Faculty Modeling
4. Course contents & delivery improvement
5. Identify & minimize mismatch between programs with market demands

With Big Data technology, we eliminate the pre-processing of collected data such as moving, cleaning, and interpolating. With streaming data and automated data management technology, we can focus on algorithmic designing and tuning, not on preprocessing. There're many algorithms available for use. The most popular are Bayes networks, association analysis/rules, classification, clustering methods, process mining, psychometric models, reinforcement learning, sequencing, space-searching algorithms, statistical analysis, and visualizations.

(The End)