Majd Sakr
Teaching Professor
Computer Science Department
School of Computer Science

TEACHING & LEARNING SUMMIT 2016

Mrigesh Kalvani
Graduate Student, MCDS/SCS

Cameron Dashti
Research Programmer, CSD/SCS

Carnegie Mellon University

Managing Cloud Resource Cost for a Large Online Project-Based Course

- Analyze sources of spending when teaching a large global course on the cloud.
- Explore different cost-saving techniques over eight semesters.
- Focus on reducing cost-per-student to create a sustainable model for large courses.

Course Overview

- Online course on cloud computing offered to Carnegie Mellon's global students.
- One of our main objectives is to give students a real-world experience.
- Our guiding principle for this hands-on project-based course is:
  - **real-world** projects, datasets, and cloud infrastructure within a budget.

Problem Statement:

- Using commercial infrastructure for student projects could lead to:
  - Overspending due to student oversight.
  - Careless release of credentials leading to account compromise.

Fig 2: Engagement Model

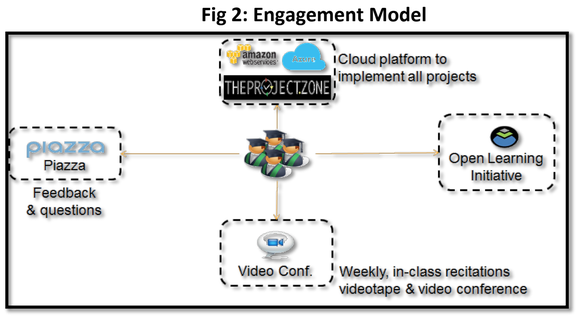


Table I: Cloud Onboarding Model

Coupon Model	Cloud Account Linking Model
Students track expenses	Both students and staff can track
No global view into usage	Can build central management
More prone to human error	Student mistakes can be identified
Cannot revoke access	Can revoke access

Fig 3: AWS Cloud Account Linking Process

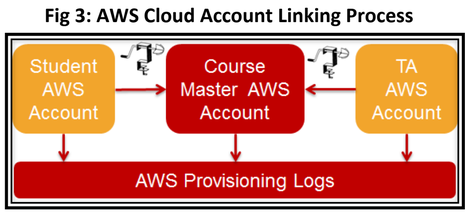
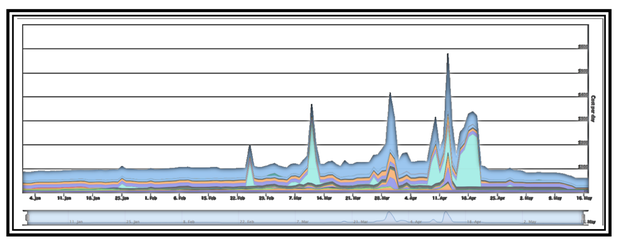


Fig 4: Our view of AWS Cloud Expenditure over a semester



Method

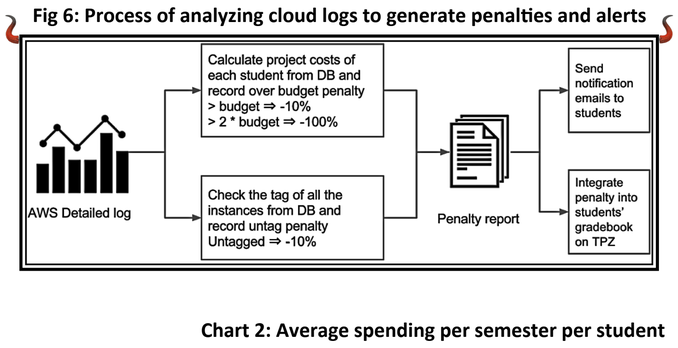
To reduce overspending, we monitor all student expenditure logs and provide access to students. We apply a penalty if a student overspends, 10% if > budget; 100% if > 2xbudget, using the following daily process:

- Parse all logs (AWS/Azure), store in a database, display in student view on theproject.zone.
- If a student overspends, a penalty is applied, automatically emailed, and displayed on theproject.zone.
- Send a daily summary report to the course staff for analysis.

Policies per course offering

- S13, F13: Collect and analyze cost logs
- S14: Email students cost warnings
- F14: Apply penalty for overspending
- S15: Deploy Netflix ICE for monitoring
- F15: Require tagging and apply tagging penalties
- S16: Monitor expenditure of course staff
- F16: Warning emails and alerts to course staff

Fig 6: Process of analyzing cloud logs to generate penalties and alerts



Evaluation

Chart 2: Average spending per semester per student

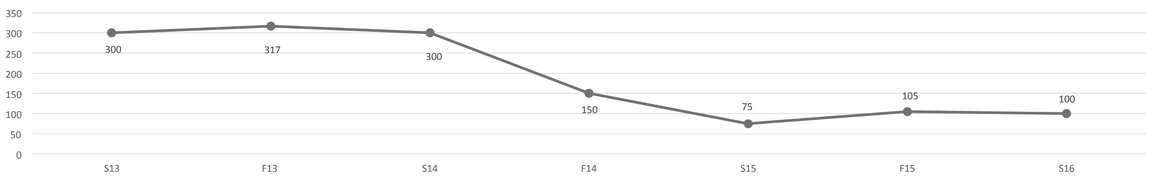


Chart 3: Overall expense/semester

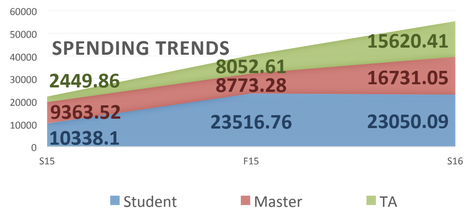


Chart 4: Student costs reduce while TAs spend more

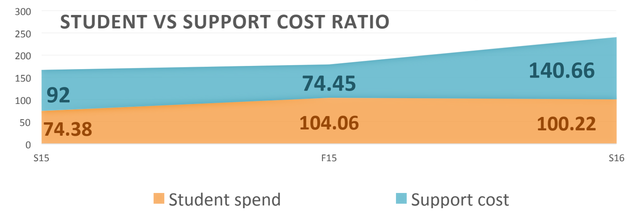


Chart 1 : Enrollment in the course

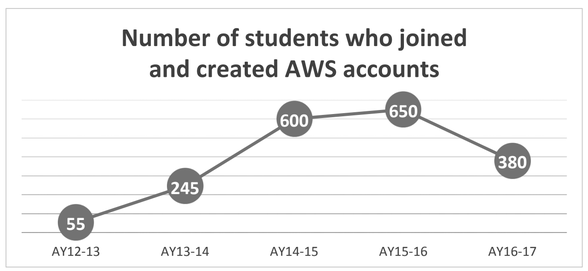
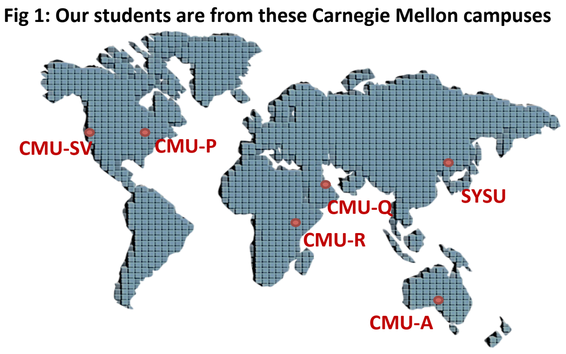


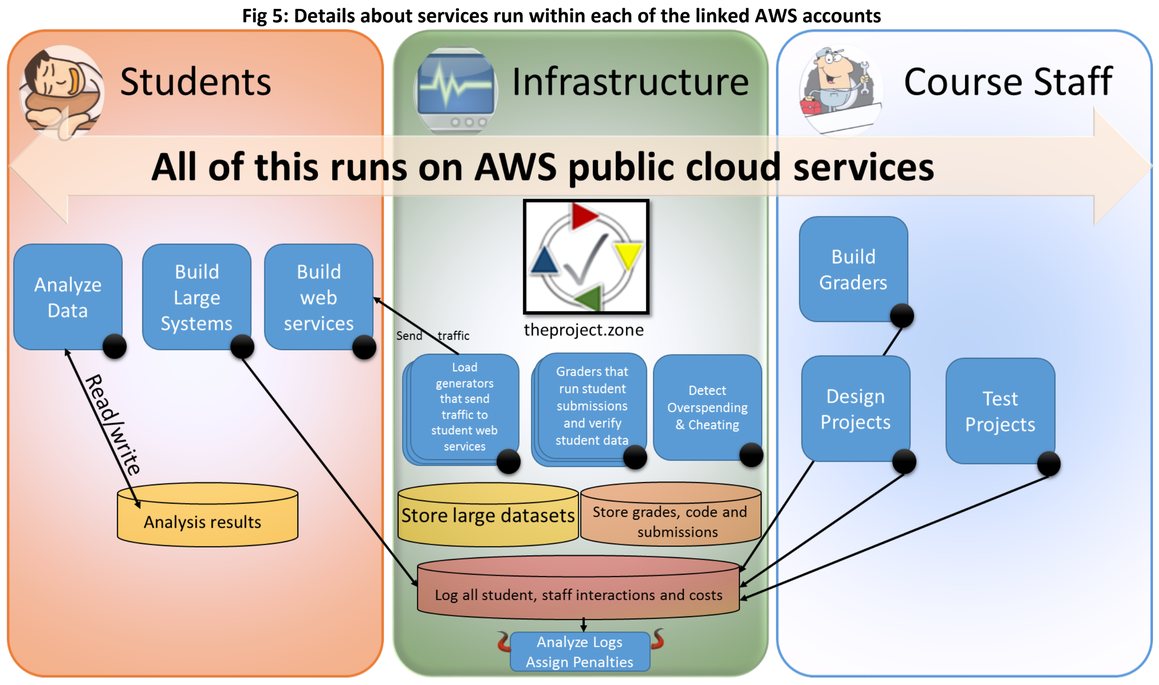
Fig 1: Our students are from these Carnegie Mellon campuses



Typical Project Design and Deployment:

Students solve projects by building large distributed systems that run on public cloud resources within an assigned budget. These systems are evaluated by autograders on the cloud that either run the student code or send HTTP web traffic to their deployed services. The blue boxes are primarily virtual machines (compute), the cylinders are storage services. Each service write logs (●) to persistent storage, which are saved in the master account. These logs are periodically analyzed for budget violations.

Fig 5: Details about services run within each of the linked AWS accounts



Lessons Learned

- Grade penalties are an effective mechanism to reduce student spending on the cloud.
- Students continue to try to bypass spending limits. We control the ability of students to do project work using personal (unlinked, unlogged) accounts.
- Cloud log data is delayed and inconsistent. Detailed monitoring requires accurate analysis to reduce false positives.
- Enforcing cloud resource tagging improves visibility at the project level.
- Students' AWS accounts continue to get hacked. Teach best practices, detect compromised accounts in real-time and consider grade penalties.
- Project quality implies testing and increases cost. Consequently, average TA spending has been increasing per year. Additionally, TAs cannot be penalized and are more likely to have runaway costs.
- Course data and cost accumulate over time, archiving is necessary.
- Each public cloud produces distinctive logs and needs specific parsing.

Table 2: Breakup of cost by resource type

Table I	Compute	Storage	Network	Misc
AY14-15	63%	18%	8%	11%
AY15-16	48%	27%	5%	20%

Table 3: Data tends to accumulate

Table II	Amount of Data	Monthly Cost
AY14-15	6 TB	$250
AY15-16	9 TB	$350

Carnegie Mellon University School of Computer Science · Open Learning Initiative · THEPROJECTZONE · Microsoft Azure · amazon web services™