Case Study on Migration of Computer Experiment Teaching Environment to Amazon Cloud Platform

Wenhan Zhan[1], Yazhu Xu[2*], Wenyu Chen[3], Hualong Huang[4]

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, China

[1]zhanwenhan@uestc.edu.cn, [2*]xuyazhu@uestc.edu.cn
[3]cwy@uestc.edu.cn, [4]hlhuang_uestc@163.com

Abstract. The concept of "Emerging Engineering Education" and "Outcome-Based Education" have put forward new requirements for the experimental teaching environment of computer discipline. Cloud computing boasts abundant resources, high elasticity, easy operation and maintenance, as well as safety and reliability. It effectively addresses the limitations of traditional on-premise labs, such as limited resources, low utilization rates, poor flexibility, and high maintenance costs. Taking the course "Comprehensive Experiment on Big Data Storage and Management" as an example, this paper introduces the practical process of migrating the experimental environment of the course from the traditional computer labs to the Amazon cloud platform, summarizes the advantages and innovations of the migration case, and provides a valuable reference and guidance for the practice of computer experimental courses on the cloud.

Keywords: Cloud computing, Experimental teaching, Experimental environment, Big data storage and management.

1 Introduction

As information technology (IT) has developed rapidly, computers have become an indispensable and essential tool in modern society, with their applications becoming increasingly widespread. As an applied discipline, experimental and practical teaching is paramount to computer education. It can deepen students' understanding of theoretical knowledge and improve students' engineering practical abilities and problem-solving skills, better meeting society's demand for high-quality computer talent. Against this backdrop, domestic universities have responded to the "Emerging Engineering Education" initiative and "Outcome-Based Education" by emphasizing computer experimental teaching and improving the quality of computer education [1][2].

The traditional lab-based computer teaching environments have problems such as limited experimental equipment, waste of teaching resources, and difficulty in maintaining. These issues not only impact teachers' teaching quality but also hinder personalized guidance and training for students. On the other hand, cloud computing platforms offer abundant resources, high elasticity, 24/7 availability, and ease of deployment and

L. Chang et al. (eds.), *Proceedings of the 2024 8th International Seminar on Education, Management and Social Sciences (ISEMSS 2024)*, Advances in Social Science, Education and Humanities Research 867,
https://doi.org/10.2991/978-2-38476-297-2_103

management, making them an excellent support system for computer experiment teaching and meeting the requirements of "Outcome-Based Education." Therefore, building a cloud-based computer experimental environment is of great significance [3][4].

As a demonstration unit for the "Emerging Engineering Education" initiative and "Outcome-Based Education," the School of Computer Science and Engineering at the University of Electronic Science and Technology of China (UESTC) has deeply collaborated with Amazon Web Services (AWS), Inc. and successfully migrated the experimental environments of several computer courses to the AWS cloud platform, achieving excellent teaching effects. Taking the experimental course "Comprehensive Experiment on Big Data Storage and Management" as an example, this paper introduces the practical process of migrating the course from traditional lab-based environments to the AWS cloud platform, highlights the advantages and innovations of this case, and provides valuable reference and guidance for the cloud practice of computer experimental courses.

2 The Current Situation and Problems of Lab-based Computer Teaching Environment

Traditional computer experiments are mainly carried out in on-premise computer labs, and the unified physical platform provides a stable and reliable experimental environment for teaching. However, there have always been significant problems in the traditional lab-based environment.

2.1 Hardware Resource Limitations

Traditional lab-based teaching environments are limited by physical space, hardware configuration, and equipment quantity, making it difficult to meet the experimental equipment needs of different courses. due to the rapid update and replacement of computer equipment, most universities are unable to keep up with the timely renewal, which fails to meet the teaching requirements.

2.2 Waste of Teaching Resources

Traditional lab-based environments require the purchase of a significant number of computer and server equipment. However, due to factors such as security management and classroom scheduling, these devices are often left idle outside of class times, resulting in an overall low utilization rate and significant resource waste.

2.3 High Maintenance Costs

The equipment in traditional computer laboratories requires regular maintenance and updates, including operating system upgrades, teaching software installation, and patch updates. These tasks are often tedious and time-consuming. Furthermore, unexpected hardware failures are common, which may disrupt the teaching process and even delay the teaching schedule.

2.4 Problems on Experimental Environment Reliability

Due to factors such as procurement and maintenance, the equipment in traditional computer laboratories often exhibits differences in hardware and software configuration. This lack of consistency in the experimental environment challenges the reproducibility of experimental results, adding complexity and uncertainty to the teaching process.

2.5 Difficulty Meeting Specialized Needs

Issues such as fixed open hours, limited available software, and relatively low user permissions exist in traditional computer lab settings, making it challenging to meet teachers' and students' personalized and flexible needs during the experimental teaching process. In recent years, the widespread adoption of emerging teaching paradigms, such as massive open online courses (MOOCs) and small private online courses (SPOCs), has further rendered the traditional computer lab environment obsolete.

It is evident that the traditional lab-based teaching environment can no longer guarantee further improvement in the quality and effectiveness of computer teaching. In the new era of computer education, new technologies and methods must be introduced to meet the requirements of experimental teaching environments.

3 Cloud-based Computer Lab Environment

Cloud computing is a modern computing paradigm based on the Internet. It provides computing resources, including hardware, software, and data, to users as services over the network to meet various IT resource requirements. Building computer lab environments based on cloud computing technology offers numerous significant advantages:

3.1 On-demand Resource Allocation

Cloud computing platforms are built and operated using a large-scale shared economy model, providing vast computing, storage, and networking resources without resource limitations. Teachers and students can easily access various software and hardware resources as needed before conducting experiments and release them after completing the experiments, thus avoiding any issues of resource wastage.

3.2 Extremely High Flexibility

Based on virtualization technology, the hardware and software experimental environments provided by cloud platforms can be highly configurable and completely unified. They can meet the requirements of the majority of computer lab courses while also reducing the uncertainties caused by the inconsistency of the lab environment. The advantages of on-demand allocation and network access also make remote experiments for E-learning possible.

3.3 Convenient Management and Maintenance

Cloud platforms enable automated management and monitoring of various virtualized resources. They also provide multiple layers of security measures, ensuring high management efficiency, reliability, and security. Teachers and administrators only need to manage and maintain the experimental environments remotely, significantly reducing labor costs.

3.4 Easy Integration with Emerging Teaching Paradigm

Cloud-based computer lab environments offer high flexibility, remote accessibility, automation, and ease of maintenance. They also feature convenient secondary development capabilities, making integrating with existing learning management systems (LMS) easier [5]. These characteristics can effectively combine with emerging teaching paradigms such as MOOC and SPOC, efficiently empowering these new teaching methods.

3.5 Low Construction and Usage Costs

Cloud-based lab environments can significantly reduce construction and maintenance costs compared with traditional labs. The experimental environments are deployed on-demand in the cloud, requiring only low-performance remote access terminals on-premise, greatly reducing capital expenditures. Additionally, given the lab equipment's lifespan, the cost of using cloud resources on a pay-as-you-go basis is typically much lower than the costs associated with purchasing, upgrading, and maintaining physical lab equipment.

Considering the aforementioned advantages of cloud-based computer lab environments, many universities and educational institutions have started migrating their lab teaching environments to the cloud.

4 Experimental Environment Migration and Optimization Case Based on the AWS Cloud Platform

Amazon is currently the largest cloud service provider globally, and AWS is the leading cloud computing service platform in market share. AWS offers robust support and assurance for various scenarios, making it the go-to platform for a wide range of industries and applications. Compared to other cloud platforms, AWS has distinct advantages in terms of stability, flexibility, security, and cost-effectiveness [6]. Choosing AWS as the underlying platform to build computer lab environments provides strong support for experiments and improves the efficiency and quality of teaching.

The School of Computer Science and Engineering at UESTC has a long-standing and deep collaboration with AWS in talent development, having successfully migrated the lab environments for multiple computer courses to the AWS cloud platform. In addition, AWS also leverages its AWS Academy and AWS Educate platforms to provide abundant free cloud skills training and cloud resource credits to teachers and students who migrate their courses to the cloud. Taking the course "Comprehensive Experiment on Big Data Storage and Management" as an example, this paper introduces the practical process of migrating its lab environment to the cloud and summarizes the advantages and innovations of this migration case.

4.1 Introduction to the Course

"Comprehensive Experiment on Big Data Storage and Management" is a practical mandatory course offered to senior students majoring in Big Data Science at UESTC. The course aims to provide students with hands-on experience on the mainstream big data processing platform, Hadoop, to help students understand the fundamental principles of big data storage, processing, and analysis, as well as master relevant skills and methods for big data processing. The course is designed in a progressive manner, starting with easy concepts and gradually progressing to more advanced topics. The main content of the course is listed in Table 1.

4.2 Introduction to the Course

The course begins with a series of fundamental experiments on the Linux operation system and key Hadoop components, where students gain initial experience with the experiment environment and the Hadoop ecosystem. Next, students engage in "MapReduce Practice" to grasp the execution process and programming methods of distributed data processing. Subsequently, they delve into "Data Warehouse Practice," which involves utilizing HiveSQL to analyze and process large-scale data [7]. Finally, students participate in a "Data Analysis Project" where they analyze data on a topic of interest and present their findings through analysis reports and technical reports [8].

Table 1. The Main Content of "Comprehensive Experiment on Big Data Storage and Management".

Experiment	Content
A. Linux Hands-on	1. Basic Command Line
	2. Filesystem Management
	3. Network Management
	4. Software Installation
B. HDFS Experience	1. Hadoop Installation
	2. HDFS Command Line
	3. HDFS Programming
C. ZooKeeper Experience	1. ZooKeeper Installation
	2. ZooKeeper Command Line

Experiment	Content
	3. ZooKeeper Programming
D. Hbase Experience	1. Hbase Installation 2. Hbase Command Line 3. Hbase Programming
E. MapReduce Practice	1. Yarn Basic 2. Word Count Practice
	3. Table selecting
F. Data Warehouse Practice	1. Hive Installation
	2. HiveSQL Practice 3. Extract, Transform, and Load
G. Data Analysis Project	Choose a big data-related topic and perform corresponding data analysis.

4.3 Experimental Environment in its Original State

Before migrating to the cloud, the lab sessions for this course were conducted in a traditional computer lab environment. However, the actual configuration of the lab's hardware, software, and network environment posed significant obstacles to the smooth delivery of the lab experiments.

Limitations on Software Environments.

The setup environment for the Hadoop big data processing platform is based on the Linux operating system and relies on numerous third-party software components. However, traditional lab computers predominantly use the Windows operating system. Although the lab computers are usually equipped with virtualization software, an additional preparatory step for the lab sessions was required: the course instructor had to create a unified virtual machine image and copy it to each physical machine to ensure that every student had a usable and consistent lab environment. This preparation is always time-consuming and error-prone.

Limitations on Lab Restoration Systems.

The lab experiments for this course are designed progressively, where each subsequent experiment relies strictly on the environment created by the preceding one. As a result, the lab environment needs to be preserved after each experiment. However, in order to ensure software security, computers in traditional lab-based environments are often equipped with system restoration features, which erase the lab environment once the machines shut down. Consequently, students had to carry portable hard drives and repeatedly copy their lab environments based on the virtual machine image. Any slight mistake could result in the inability to proceed with the subsequent experiments.

Limitations on the Hardware Configuration.

To ensure the utmost resemblance to production environments, a five-node Hadoop cluster was deployed for the experiment. Nevertheless, due to limited computer lab hardware configurations, running five virtual machines simultaneously on a single physical machine and executing resource-intensive data mining tasks becomes challenging. Consequently, the instructor had to resort to running only one virtual machine per physical machine and implement a cluster setup by networking these virtual machines. This approach further adds to the complexity of the experiment's configuration.

Limitations on Network Environments.

Due to the cluster setup involving virtual machines across multiple physical machines, the virtual and physical machines must share the same network. This tight coupling between the network configuration of the virtual machine cluster and the physical network configuration poses challenges. In computer labs where IP addresses are fixed, a brute-force solution can be implemented by assigning specific seats to students. However, in labs where IP addresses are dynamically allocated (DHCP), the cluster's IP addresses must be reconfigured before each class session. This process is complex, prone to errors, and can significantly impact subsequent experiments. Additionally, since all machines in the lab are in the same local area network, multiple Hadoop clusters built by students may interfere with each other. Misconfigurations in one cluster can propagate to others, making troubleshooting and resolving such issues highly complex.

Data Sharing Problem.

Due to the complexity of this course, instructors and students often need to frequently share large files such as virtual machine images and datasets for analysis. These files are typically huge, with single files usually exceeding 10GB. However, in the computer lab environment, achieving fast file sharing is challenging due to bandwidth and machine performance limitations, wasting valuable class time as the file transfer process becomes time-consuming and inefficient.

The aforementioned obstacles contribute to a significantly high level of complexity in teaching the “Comprehensive Experiment on Big Data Storage and Management” course. Typically, instructors require extensive preparation before each class session. Additionally, multiple graduate teaching assistants are needed to promptly address various environmental and configuration issues that arise during the lab experiments. Despite these efforts, many problems remain unsolved during the teaching process, resulting in many students being unable to complete all the experiments and providing poor evaluations of the course.

4.4 Migration of the Lab Environment

By migrating the lab environment of this course to the AWS cloud platform, the problems mentioned above have been effectively resolved, significantly reducing the complexity of teaching and the difficulty of learning. Below is a brief overview of the AWS cloud services required to move the experimental environment to the cloud.

Amazon EC2.
AWS offers cloud-based virtual servers through Amazon Elastic Compute Cloud (EC2). Similar to local virtual machines, EC2 instances provide features such as image creation, snapshots, and virtual disks (through Amazon EBS service). Additionally, EC2 instances excel in terms of performance, security, and stability.

Amazon VPC.
Through the virtual private cloud (VPC) service, users can create an isolated virtual network within the AWS cloud environment. This service allows users to control inbound and outbound network traffic rules, enabling network traffic isolation and secure management. At the same time, EC2 in the VPC can have a fixed public IP address (through Amazon Elastic IP service) and occupy it for a long time.

Amazon S3.
Amazon Simple Storage Service (S3) provides an object storage service that enables high-speed uploading, downloading, and sharing of files of various types and sizes. This service offers features such as high performance, scalability, and reliability. It allows for efficient storage and retrieval of objects, making it suitable for a wide range of use cases.

AWS CloudWatch.
AWS CloudWatch is an essential monitoring and observability service that enables users to gain insights into the performance of various AWS resources and applications, track metrics, analyze log files, and receive alarm notifications. It provides a centralized platform for monitoring and troubleshooting the cloud infrastructure.

AWS CloudFormation.
AWS CloudFormation is a robust and scalable infrastructure-as-code service that enables the description of infrastructure requirements in a declarative template script and automates the provisioning and management of AWS resources. It makes creating, updating, and deleting stacks of cloud resources easier, ensuring consistency and repeatability in users' deployments.

Based on the aforementioned AWS services, it is extremely convenient to build the lab environment for this course in the cloud. First, the Amazon VPC service is adopted to create an isolated virtual private network in the cloud. All cloud resources launched within this network are part of the same local area network, eliminating interference from other networks and resolving the mutual influence issues caused by multiple clusters sharing the same network in the lab-based on-premises environment. Then, the required number of EC2 instances are created within the VPC. The virtualized cloud environment ensures that the hardware and operating system environments of all EC2 instances are completely consistent and highly configurable, efficiently resolving the software and software configuration limitations encountered in traditional labs. Furthermore, each EC2 instance can be assigned a public IP address through the AWS Elastic IP service, facilitating remote access for students to conduct their experiments.

After the above virtual infrastructure is successfully configured, students can progressively perform the experiments, with the operation process being mainly similar to that of a local lab environment. During the experiments, students can create snapshots of their EC2 instances as needed, allowing them to quickly restore the previous experimental environment in case of configuration issues. Moreover, the EC2 snapshots are directly stored in the cloud, eliminating the need for data transfers using portable storage devices. All large-sized data required for the experiments, such as public EC2 snapshots and big data sets, are stored in Amazon S3, enabling convenient cloud data sharing between the instructor and students.

After completing each experiment session, the students only need to stop the running EC2 instances. Thanks to the Amazon EBS service, the data within EC2 instances is preserved even after they are stopped. When the students need to continue their experiments, they can simply start the EC2 instances again, and the previous experimental environment can be restored without being hindered by the system restoration issues commonly encountered in traditional on-premises labs.

4.5 Optimization of the Cloud-based Lab Environment

Through the migration of the experimental environment as described above, the successful and effective implementation of the "Comprehensive Experiment on Big Data Storage and Management" course is already ensured. However, to further reduce the difficulty of getting started with this experimental environment, improve classroom efficiency, and effectively integrate it with emerging teaching paradigms, such as MOOC and SPOC, the course team has also carried out the following optimizations to the cloud-based experimental environment:

Multi-stage EC2 Images.

As mentioned earlier, this course adopts a progressive experimental teaching approach, where subsequent experiments strictly depend on the success of the previous experiments. However, in teaching practice, it is common for some students to be unable to complete a certain experimental stage, which would then impact the normal progression of all subsequent experiments. To address this problem, in addition to the initial EC2 image, the course team has prepared a series of EC2 images for each experimental stage. This way, even if issues arise in a particular experimental phase, students can directly rebuild the subsequent experimental environment using the pre-prepared multi-stage EC2 images, allowing them to temporarily skip the current problematic step without affecting their overall learning progress.

Experimental Environments as Code.

Rebuilding the initial experimental environment based on the multi-stage EC2 images is relatively complex (requiring the reconstruction of the VPC, EC2 instances, and extensive configuration of the experimental environment). It is not user-friendly for instructors and students lacking cloud platform operations experience. To address this, the course team has created resource template scripts with the infrastructure-as-code service, i.e., AWS CloudFormation, for each experimental stage of the course. This way, the instructor can run the corresponding script whenever needed to recreate the experimental environment, significantly reducing the complexity.

Integrating with Amazon Learning Management System (LMS).

The migration and optimization measures mentioned above have successfully addressed the on-premises experimental requirements of this course. However, the cloud-based environment still faces two challenges: 1) the relatively high usage costs for cloud resources and 2) the difficulty in supporting emerging teaching paradigms, such as MOOC and SPOC. To address these issues, with Amazon's strong support, the course team has integrated the experimental environment with AWS Academy's LMS. Through the LMS, the instructors can allocate a virtual AWS account to each student, with each student account including a 100 free usage credit for AWS cloud resources. Students can independently conduct the experiments on the AWS platform with their assigned accounts. With Amazon's support, the course team has successfully resolved the problem of high cloud usage costs. Furthermore, with the help of this LMS, the course experiments are no longer constrained by geographical location. As long as the network is accessible, students can perform the course experiments at any time and location, effectively supporting emerging teaching paradigms.

4.6 Teaching Effectiveness and Evaluation

Since migrating to the cloud, the experimental course has received a positive response from both students and teachers. Students generally find the cloud-based experimental environment to be more convenient, stable, and free of hassle, enabling personalized learning and ensuring the smooth progress of the experiments. For teachers, the migration has brought a sense of liberation: On the AWS cloud platform, the management and maintenance of the experimental environment have become much simpler and more efficient, making remote teaching possible. These positive feedback points underline the cloud platform's value and advantages in experimental teaching.

Influenced by the success of this course, the School of Computer Science and Engineering at UESTC has now completed the migration of the experimental environments for five additional courses to the cloud, including “Cloud Computing Fundamentals and Practice,” “UNIX Operating System,” and “Software Development Comprehensive

Experiment." These cloud-based experimental environments have all achieved excellent teaching results, and the process of migrating experimental environments to the cloud is expected to accelerate further.

5 Conclusion

Compared to the traditional lab-based environment, the cloud-based computer experimental environment offers numerous advantages. By migrating the experimental environment to the AWS cloud, universities can provide students with a superior learning experience, enhance overall teaching effectiveness, and proactively address the evolving demands of modern education.

Acknowledgments

This work was supported in part by the Experimental Teaching and Lab Construction Research Project of the University of Electronic Science and Technology of China under Grant BK-SYJY-202410 and in part by the National Demonstration Center for Experimental Computer Education of the University of Electronic Science and Technology of China.

References

1. Attila Bencsik, Krisztina Némethy, Gyula Kártyás, and Franciska Hegyesi. 2015. Effects of new engineering paradigm on university education programs. In Proceedings of the 2015 IEEE 10th Jubilee International Symposium on Applied Computational Intelligence and Informatics. IEEE, Timisoara, Romania, 465-469. doi:10.1109/SACI.2015.7208249.
2. Xiaoyan Wu, Shu Wang, and Min Xu. 2021. Exploration of Experiment Teaching of Mechanical Specialty Based on New Engineering Background. In Proceedings of the 2021 International Conference on Computers, Information Processing and Advanced Education (CIPAE). IEEE, Ottawa, ON, Canada, 334-337. doi: 10.1109/CIPAE53742.2021.00086.
3. Hana Eljak, Ashraf O. Ibrahim, Fakhreldin Saeed, Ibrahim A. T. Hashem, Abdelzahir Abdelmaboud, Hassan J. Syed, Anas W. A., Mohd Arfian Bin Ismail, and Abubakar Elsafi. 2024. E-Learning-Based Cloud Computing Environment: A Systematic Review, Challenges, and Opportunities. IEEE Access 12, 7329-7355.
4. Meirong Lin, and Jie Chen. 2023. Practice teaching exploration of database system based on huawei cloud. In Proceedings of the 2023 International Conference on Computer Engineering and Distance Learning (CEDL). IEEE, Shanghai, China, 41-45. doi: 10. 1109/ CEDL60560.2023.00015.
5. Sebastian Kucharski, Iris Braun, and Tommy Kubica. 2023. An Adaptive, Structure-Aware Intelligent Tutoring System for Learning Management Systems. In Proceedings of the 2023 IEEE International Conference on Advanced Learning Technologies (ICALT) IEEE, Orem, UT, USA, 367-369. doi: 10.1109/ICALT58122.2023.00113.
6. Yimeng Cai, Yuchen Luo, Yijun Qiu, and Xinning Zhao. 2022. The adoption and deployment of cloud computing: a comparative study between China and the US. In Proceedings

of the 2021 International Conference on Big Data, Artificial Intelligence and Risk Management (ICBAR). IEEE, Shanghai, China, 168-174. doi: 10.1109/ICBAR55169.2021.00042.
7. Bhola Gan Chaudhuri, and Shalli Rani. 2024. Managing Metadata in Data Warehouse for Data Quality and Data Stewardship in Telecom Industry - A Compact Survey. In Proceedings of the 2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). IEEE, Greater Noida, India, 353-361. doi: 10. 1109/ ICCCIS60361. 2023.10425001.
8. Chabane Djeddi, Nacer E. Zarour, and Pierre-Jean Charrel. 2022. A Novel Method of Requirements Analysis for Big Data Projects. In Proceedings of the 2022 4th International Conference on Pattern Analysis and Intelligent Systems (PAIS). IEEE, Oum El Bouaghi, Algeria, 1-7. doi: 10.1109/PAIS56586.2022.9946665.