CUTE Labs: Low-Cost Open-Source Instructional Laboratories for Cloud Computing Education

Dr. Keke Chen, Wright State University

Keke Chen is an associate professor in the Department of Computer Science and Engineering, a member of the Ohio Center of Excellence in Knowledge-Enabled Computing (the Kno.e.sis Center), at Wright State University. He directs the Data Intensive Analysis and Computing (DIAC) Lab at the Kno.e.sis Center. He earned his Ph.D. degree from Georgia Institute of Technology in 2006, his Master's degree from Zhejiang University in China in 1999, and his Bachelor's degree from Tongji University in China in 1996. All degrees are in Computer Science. His current research areas include cloud computing, secure data services and mining of outsourced data, the privacy issues in social computing, and visual exploration of big data. During 2006-2008, he was a senior research scientist at Yahoo! Labs, working on web search ranking, cross-domain ranking, and web-scale data mining. He owns three patents for his work in Yahoo!.

Prof. Bin Wang, Wright State University

Prof. Bin Wang earned his Ph.D. from the Ohio State University in 2000. He joined the Wright State University in September 2000, where he is currently full professor of computer science and engineering. His research interests include optical networks, real-time computing, mobile and wireless networks, cognitive radio networks, trust and information security, and semantic web. He is a recipient of the US Department of Energy Career Award. His research has been supported by US Department of Energy, National Science Foundation, Air Force Office of Scientific Research, Air Force Research Laboratories, Ohio Supercomputer Center, and the State of Ohio.

Prof. Prabhaker Mateti, Wright State University

Prabhaker Mateti, Ph.D. in Computer Science, 1976, University of Illinois at Urbana-Champaign.

My research interests are in Internet security, distributed systems, programming language design, technical aspects of software engineering, and graph algorithms. My recent work is aimed at strengthening the security of Operating Systems and the Internet via auditing the existing code with the aid of mathematical verification tools, and redesigning with security as the primary goal. I regularly teach, among others, a course on Security that was developed with funding from NSF.

CUTE Labs: Low-Cost Open-Source Instructional Laboratories for Cloud Computing Education

Abstract

Compared to the fast development of cloud-based applications and technology, higher education on cloud computing is seriously lagging behind. Built upon our experience on cloud computing education in the past years, a set of laboratories for Cloud compUTing Education (referred as the CUTE labs) has been developed and evaluated. The CUTE labs cover four major areas: the platform exploration labs, the big data labs, the cloud economics labs, and the security and privacy labs. They are designed to use publicly available free cloud resources and open source software with no special requirement on computing infrastructures, so that they can be easily adopted and adapted at low cost. Currently, we have developed thirteen labs, among which eleven have been deployed during 2009-2015 and six have been evaluated during 2014-2015. The preliminary results show that these labs have successfully reached their initial design goals.

1 Introduction

Cloud computing has become a new popular computing paradigm. It has shown great potential in business, research, and education. With the pay-as-you-use utility computing model and virtualization techniques[3], cloud computing can significantly reduce the cost of computing and easily scale up or down upon users' requests. In fact, many popular online computing tools and services that we enjoy today are built on clouds, such as Dropbox, Facebook, and most of Google's products.

With cloud computing, it becomes possible for users to provision a large amount of computing power in a short time to perform large-scale data intensive computation, and to be charged only for the time and resources used This unique feature has attracted researchers in scientific computing[27,28,30] and data mining[25,19] to explore the potential of cloud computing in solving big data problems. Cloud computing will also bring a revolution to education[29,14,18,26]. In particular, cloud computing can reduce the IT cost in education, and level the playing-field of computing infrastructures for cash-strapped institutions.

However, the education on cloud computing is seriously lagging behind the fast development of the cloud computing technology. Recent independent survey[13] still shows that a large number of respondents have little knowledge about cloud computing. The lack of knowledge also makes them hesitate to adopt the technology. We have seen that a large demand on cloud computing professionals has emerged[22]. A gap is forming between the fast growing cloud computing market and the scarcity of qualified cloud computing professionals. We have to ask ourselves: *how to fill this gap.*

Cloud computing education, like the traditional topics in computer science, such as Operating Systems, Compilers, Databases, and Networks, needs effective hands-on laboratory exercises to support lectures. The importance of experiential learning has long been recognized in the learning theory literature[21]. We should engage students in hands-on experience in cloud computing education, and thus well-designed laboratory exercises are critical to the success of cloud computing education. The traditional CS courses have taken a long time to establish and adopt a set of laboratory exercises. However, since cloud computing education is still embryonic, only a few publicly-accessible individual laboratories exist. Nationwide there does not exist an easily replicable model for designing and running cloud computing labs.

To fill the gap in cloud computing education, we are developing an easily adaptable laboratory environment and a comprehensive list of laboratory exercises that are essential to cloud computing education. Based on our experience in teaching cloud computing classes, we have identified four major aspects in cloud computing education: *cloud systems*, *data intensive scalable computing*, *cloud economics*, and *cloud security and privacy*. During the past years we have developed more than ten labs to cover these four aspects, and this collection is still growing. These labs work together to help students learn the cloud computing principles and engineering skills. We name this set of labs: the CUTE labs (Cloud compUTing Education labs)[1]. They have been experimented and evaluated in courses such as Cloud Computing, Operating Systems Internals, Internet Security, Distributed Systems, and Privacy-Aware Computing for both undergraduate and graduate students at Wright State University during the past two years.

CUTE labs have a number of appealing properties.

- CUTE labs cover a wide spectrum of cloud computing concepts and principles, as well as important engineering skills. A general software/hardware framework is designed to support sustainable lab development.
- CUTE labs are low-cost labs, and can be implemented with free cloud resources (e.g., Amazon Web Services in Education and Google App Engine) and open source software. They are replicable by even the most cash-strapped institutions.
- CUTE labs will be open sourced and allow instructors to easily adopt, or revise to create their own labs, which will foster a community of cloud computing educational labs.

The remaining sections of the paper are organized as follows. We will first describe the related work on cloud computing education and laboratory development in Section 2. Then, we describe the design principles in Section 3 and give design details about the specific labs in Section 4. We will also report current deployment status and show some evaluation results in Section 5.

---

[1]http://www.cs.wright.edu/ keke.chen/nsf/cute.htm

2 Related Work

State of the Art in Cloud Computing Labs Cloud computing introduces a set of new concepts and technologies, which are distinct from the traditional ones in operating systems, networks, and distributed systems. These topics include hardware/software virtualization, data intensive scalable computing, cloud economics, and cloud security and privacy. In the United States, only a few institutions offer courses on cloud computing with limited labs, and often these labs are not publicly accessible. University of Maryland offered a cloud computing course entitled: INFM718G/ CMSC838G: Data-Intensive Information Processing Applications[23], focusing on MapReduce programming and its application on information retrieval. Johns Hopkins University offers a course on security and privacy in cloud computing[17] with no labs. Stanford University[10] and Cornell University[4] have seminar-style cloud computing courses, again with no labs. University of Washington has the cloud computing certificate program, but no labs are published. University of California at Berkeley has been applying cloud computing technology to computer science courses such as databases and distributed computing[15]. Overall, none of these courses has published a comprehensive set of labs to cover the fundamental principles of cloud computing. As more and more universities start or plan to offer cloud computing courses, there is a critical need to design a set of comprehensive labs that can be easily adopted, adapted, and deployed by various institutions.

Preliminary Work at WSU We have offered CS499/699: Cloud Computing[5] every year during 2009-2012. After 2012, this course was split into two courses: CEG4360 Distributed Systems and Cloud Computing, which focuses on fundamental principles of distributed computing and the basic concepts in cloud computing; CEG7380 Cloud Computing, which is a graduate-level course, including more advanced topics in cloud computing. These courses cover data intensive computing in the cloud (e.g., MapReduce, Pig, and Spark), cloud platforms (e.g., Amazon Cloud and Google App Engine), cloud security and privacy, and recent research topics in cloud computing. In this process, we learned the importance of well designed cloud computing labs. The first early developed four labs cover platform exploration and data intensive computing. Other recently developed ones thoroughly cover the major principles and techniques in cloud computing. We are also actively incorporating the cloud computing concepts and labs into other related courses. The virtualization concepts are introduced in OS Internals. Cloud security and privacy is an important topic in Internet Security, and Privacy Aware Computing. The Amazon EC2 labs will also be adopted in Mobile Computing. We are also working on several research topics related to cloud computing, including interactive visual exploration of data in the cloud[9], optimal resource provisioning for large scale MapReduce programs[8], and privacy preserving techniques for outsourced data[6,7].

3 CUTE Framework

The CUTE framework includes the design philosophy of our labs, the lab environment, and the procedure of lab development.

3.1 Design Philosophy

Like most computer science courses, cloud computing education should equally focus on both the fundamental principles and practical skills. Students should have opportunities to apply, integrate, and experiment with these principles and skills.

Fundamental Principles Among many aspects of cloud computing, *cloud systems*, *data intensive scalable computing*, *cloud economics*, and *cloud security&privacy* have been well-recognized by the researchers[3,1] and practitioners[2] as the major aspects in cloud computing. (1) Cloud computing platforms have a set of unique features such as elastic resource provisioning, instant scalability, and virtualization. We believe it is best for students to experiment with the real cloud systems to understand these features. (2) Cloud economics, based on the pay-as-you-go utility computing model, is the most important principle attracting the cloud users. It is thus important for students to understand how it works and how it is used in practice. (3) Benefited from cloud economics and resource elasticity, data intensive scalable computing (or big data analysis), which intermittently needs a large amount of resources for data intensive parallel processing, has emerged as a popular type of cloud application. It has introduced influential new programming and processing models such as MapReduce[11] and Spark[32]. We believe students with this area of knowledge and skills will be in high demand in the market. (4) Cloud security and privacy is also a well recognized problem in cloud computing[24]. According to a survey by Deloitte[16] on cloud computing, 30% of the surveyed companies were most worried about their own intellectual properties on clouds, and 15% worried about unauthorized use of their data. It is necessary for the students to understand the risks of hosting data and computations on top of untrusted cloud platforms and the methods for mitigating these risks.

Practice Skills We believe that students should not only learn how to explain the principles, but also learn practices to solve cloud-related problems. There are several main categories of computing practice, as mentioned by Denning[12]: (1) programming, (2) engineering systems, (3) modeling and validation, (4) innovating, and (5) applying.

Correspondingly, we provide an interpretation of the practice skills in terms of cloud computing education. (a) Programming: using cloud-specific programming model and high-level languages to build data intensive analytic applications; (b) Engineering systems: designing and constructing software systems based on cloud computing infrastructure. (c) Modeling and validation: developing cloud economics models to estimate and optimize the financial cost of a specific cloud-based application; designing experiments to validate models. (d) Innovating: creatively thinking and solving problems - it is very important when cloud computing is still at the infant stage. (e) Applying: applying cloud computing principles to solve real-world problems. We have developed laboratories that jointly provide the coverage of these five essential skills in cloud computing.

[2] ibm.com/cloud, hp.com/go/cloud, highscalability.com

3.2 CUTE Laboratory Environment

We developed four types of labs: (1) platform exploration labs, (2) large scale data processing labs, (3) cloud economics labs, and (4) cloud security and privacy labs. Together, these help the students understand the fundamental principles and attain the essential skills described in our design philosophy.

(1) The objective of the platform exploration labs is to provide students with opportunities to get familiar with the most popular public cloud systems. Students learn to use the interfaces provided by these systems to build applications; students also have chances to explore the underlying working mechanisms of the cloud infrastructure.

(2) With cloud computing, it is possible to perform scalable data intensive analysis in an economical way. The objective of large scale data processing labs is to learn the popular parallel data processing techniques that work on scalable cloud infrastructure. Students get chances to work on real parallel data processing systems and learn the techniques to deploy such systems to the public cloud.

(3) Cloud economics is a unique component that other types of distributed computing do not have. The objective of the cloud economics labs is to understand and evaluate the financial aspect of cloud computing applications.

(4) Data security and privacy are essential topics in cloud computing. Cloud computing users may have to host sensitive data in the platform - without security and privacy guarantee, many users will not use clouds. The objective of this set of labs is to let students understand the security and privacy principles in cloud computing and experiment with recently developed techniques.

Labs	Platform Exploration Labs	Data Intensive Scalable Computing Labs	Cloud Economics Labs	Security and Privacy Labs
Software	Xen, boto	Hadoop/Pig/HBase	gprof	nmap, hping
Cloud systems/ hardware	Amazon EC2, Azure Google App Engine Desktop/Laptop	Desktop/Laptop, or inhouse cluster, or Amazon EC2	Amazon EC2 Desktop/Laptop	Amazon EC2 Desktop/Laptop

Figure 1: The software/hardware stack for different CUTE labs.

To make the labs easily adoptable at different educational settings, we try to minimize the dependency on the lab infrastructure. To achieve this goal, we designed all the labs using free cloud resources and open source software. We also considered a number of options in terms of hardware configuration. Figure 1 shows the software/hardware stack for the labs. All the software tools used in the labs are open source. The two cloud platforms mentioned in Figure 1 are available for free access - Amazon offers free credits for cloud computing courses and also free usage tier for new users (aws.amazon.com/free/), so that students can use the free credits to access the Amazon Cloud in the labs; Google App Engine has free quota for each account, also sufficient for the labs;

Azure also has an education program to support Cloud-related education. For the data intensive scalable computing labs, there are three different settings, depending on the resources the class may have. (1) All the labs can be done using only desktop/laptop. Hadoop, Pig, and Spark have the option to be installed as a standalone system in a personal computer. (2) If the university has a Hadoop cluster, these labs can certainly be deployed to the cluster. (3) If there is no local cluster but the instructors like to use a cluster environment, Amazon educational EC2 can be used to establish the cluster environment. Therefore, the lab environment can be easily portable to even the most cash-strapped institutions. It provides a very low entry point for nationwide adoption and adaptation.

3.3 Procedure of Lab Development

The development of each lab follows an iterative design-deployment-evaluation procedure depicted in Figure 2. The first stage is *lab design and implementation*, in which learning goals are defined, lab tasks are designed, key issues are identified, and supporting materials are prepared.

The second stage is *deployment*, in which labs are deployed in different educational settings. Results from this stage illuminate the challenges of and opportunities for adapting our labs in diverse educational settings. The deployment stage will allow us to collect data for the next stage, as well as to obtain feedback on the design of each lab, which will be used to improve the lab design. We deploy the labs to cloud computing and a number of related courses such as distributed computing systems, computer networking, operating systems, databases, and mobile computing.

The third stage is *evaluation*, in which we evaluate how effective our approach is in enhancing student learning in cloud computing education. However, this project does not focus on developing novel evaluation tools. We will use existing tools and processes for evaluation. An advisory committee oversees the evaluation procedure.

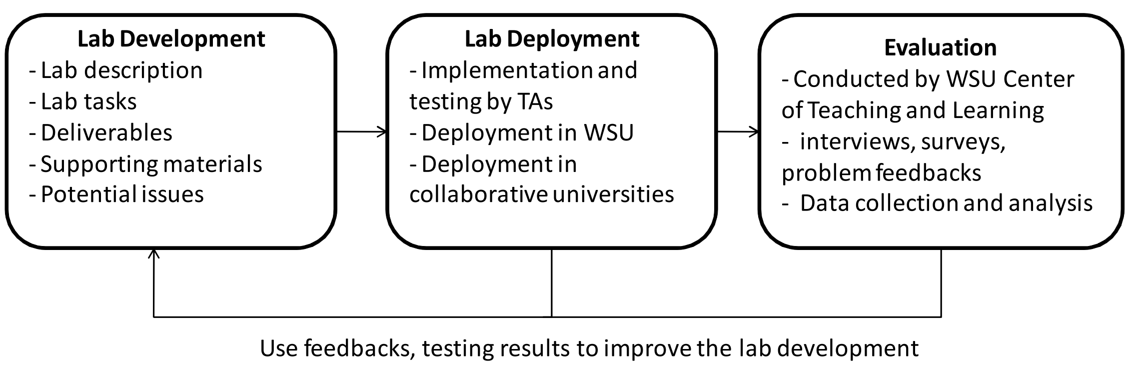

Figure 2: The procedure of lab development.

4 Design of CUTE Laboratories

In this section, we describe the four types of labs in detail. Currently we have developed 13 labs and are planning more. Instructors can select a subset to fit the needs of different courses. The specific lab descriptions can be found at our project site.

4.1 Platform Exploration Labs

An important task of cloud computing education is to teach students to use the existing cloud computing tools. The infrastructure exploration labs are designed to satisfy this requirement. In this set of labs, the students will use the most popular public cloud computing platforms such as Amazon Web Services (AWS) and Google App Engine (GAE). GAE provides a certain amount of free resources for each account. Free AWS resources are available in two forms. First, AWS provides a teaching grant for each qualified university/college course about cloud computing that uses AWS for labs. According to our previous experience, as long as the course is formally offered in a university/college with a course web site, Amazon will give the teaching grant ($100 credits per student). Second, even without the teaching grant, each new AWS account can use an amount of free credits for the first year, which will be sufficient to cover the cost of related labs.

Expected learning outcomes: Students learn to use the cloud platforms to build cloud applications; students understand the technical issues of cloud computing infrastructure.

Currently, we have developed the following labs:

- Simple AWS Lab: use commandline tools and the programming interface to manage EC2 and S3 resources.
- Advanced AWS Lab: learn how to automatically deploy and manage a cluster (e.g., a Hadoop cluster) on EC2. Learn the performance difference between S3 and instance-level local storage.
- Docker Lab: learn to use the Docker containers - a popular virtualization technique.
- Virtualization Lab: understand the differences between the OS on the bare-metal system and that on the virtualization systems.

These labs are more focused on Infrastructure-as-a-Service (IaaS). More labs will be developed on Platform-as-a-Service (PaaS), using existing popular providers such as Google App Engine and Microsoft Azure.

4.2 Big-Data Computing Labs

Recent studies[3,1] show that data-intensive applications or services (e.g., parallel batch processing[11,31] and business analytics[20]) will be the major applications in the cloud. Therefore, students who are equipped with the knowledge in the software systems and practice skills for large-scale data analytics will have good opportunities in the market. This second set of labs is designed to meet the new trends in cloud-based data analytics.

Expected learning outcomes: Students learn to use the MapReduce programming model, the Pig language, the Spark system, and cloud data store systems to solve real problems; students understand how to choose appropriate data intensive scalable computing tools for different data analytic problems.

Currently, we have developed the following labs:

- Hadoop/MapReduce Lab: get familiar with the Hadoop system and learn to program with MapReduce.
- Pig Programming Lab: learn to program with the higher-level language Pig for MapReduce processing of relational data.
- Spark Lab: learn to program with the Spark system, faster reliable parallel processing with Resilient Distributed Datasets (RDD).
- Cassandra Lab: learn to manage and program with the Cassandra key-value store.

A few more labs are under planning to address the current trend of big data processing.

4.3 Cloud Economics Labs

It is well recognized that cloud economics distinguishes cloud computing from service computing and other types of distributed computing[3]. Since cloud resources can be obtained on demand and billed only for their usage (e.g., each EC2 instance has a price per hour), one of the key problems is to minimize the cost while preserving the Quality of Service(QoS). Concretely, there are a number of related problems. (1) For Web applications hosted in the cloud, the workload is determined dynamically by users' requests. Thus, it is important to accurately *predict* the workload so that cloud resources can be provisioned *promptly* and *in appropriate size*. (2) For large-scale data analytic applications, the workload is pre-determined by the amount of data to be analyzed and the complexity of the analysis algorithm. There is an intricate relationship among the amount of resources to use, the time the job will take to finish, and the amount of money paid to the cloud provider, which can be formulated and solved as optimization problems.

Expected learning outcomes: Students understand the basic principles of cloud economics; students learn the skills for modeling and experiment with the cost models and the pricing models.

Currently, we have developed two labs:

- Resource Monitoring and Provisioning for Web Applications. Study the resource provisioning problem for Web Applications deployed in the cloud and experiment with related techniques.
- MapReduce Cost Modeling and Resource Provisioning. Study the MapReduce processing workflow and experiment with the MapReduce cost model and resource optimization methods developed by our research team[8].

We will develop one more lab about cloud pricing models.

4.4 Security and Privacy Labs

Without good understanding of cloud security and effective security and privacy-preserving solutions, many companies, government agencies and research institutes will hesitate to use public clouds. In this set of labs, we study two issues: (1) exploring the security threats that are unique to the cloud platform, and (2) experimenting with the geometric data perturbation technique[7] for privacy preserving data mining[2] in the cloud.

Expected learning outcomes: Students understand the specific security and privacy issues in cloud computing; students learn the technical skills to explore the security and privacy problems and apply the latest techniques to enhance security and privacy.

We have developed two labs:

- Data Privacy Lab. Implement and experiment with several data disguising methods for protecting data privacy in outsourced database and data mining services.
- Data Security Lab. Learn about the security issues with the cloud systems.

More security labs may be added in the near future.

5 Deployment and Evaluation

In this section, we briefly review the deployment history of the designed labs, and report the evaluation methods and results.

5.1 Deployment

We have deployed most of the designed labs to classes. Figure 3 shows the status of lab deployment. The labs are developed and deployed in three batches. The first three labs were designed and used in CS499 (Cloud Computing) and CEG435 (Distributed Computing Systems) during 2009-2012 when Wright State was still using the quarter system. Then they were continuously evolved and deployed in semester classes CEG7380 (Cloud Computing) and CEG4360 (Distributed Systems and Cloud Computing). The second batch, including four labs, was developed during 2012-2014 and used in three courses: the above two, plus CEG2350 (OS Concepts and Usage) that uses the Introduction lab. The third batch, including six labs, was developed in 2015, among which four had been deployed in CEG7380 and the other two are expected to be deployed in 2016.

5.2 Evaluation

Project evaluation is overseen by the advisory committee which consists of two renowned cloud computing researchers and educators: Prof. Jimmy Lin at University of Waterloo and Prof. Cal-

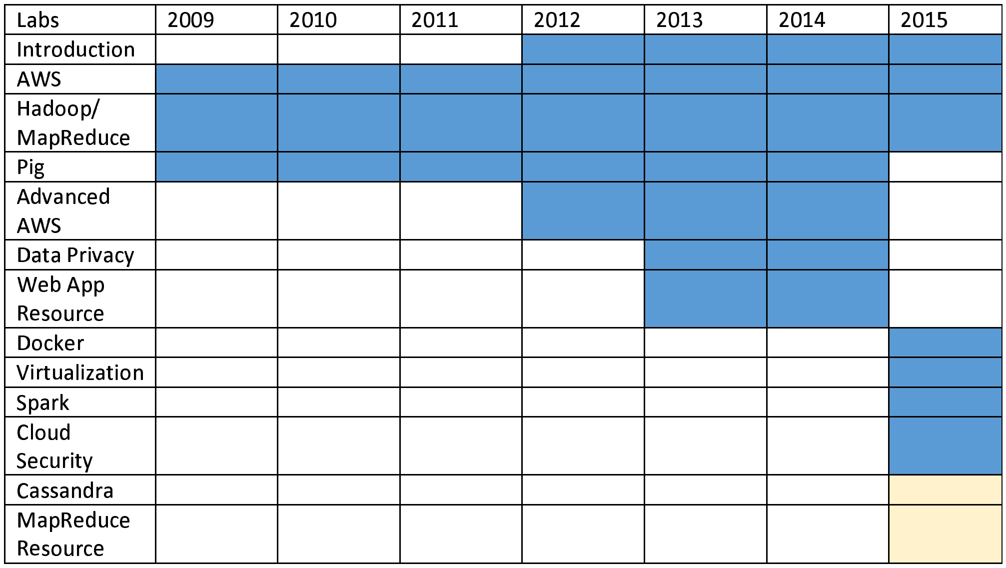

Figure 3: The deployment of current labs. Blue: deployed, Yellow: developed but not deployed yet.

ton Pu at Georgia Tech. The Center for Teaching and Learning (CTL) at Wright State provided supports for syllabus analysis and study design.

Evaluation Goals The goal of this project is to enhance student learning in cloud computing education using the proposed laboratories. The efficiency and effectiveness of the labs are the major factors to be evaluated. We quantify the lab efficiency using the following metrics: (1) the level of difficulty of a lab; (2) the usefulness of the supporting materials; (3) the time students spent on a lab; (4) whether the time spent is worthwhile; (5) whether too much time is spent on the tasks that are non-essential to the targeted concepts.

The effectiveness factor assesses student learning as a result of the project. We quantify the effectiveness using the following metrics: (1) students' level of interest in the lab exercise; (2) students' level of engagement in the lab exercise; (3) level of challenge presented by the lab exercise; (4) amount of effort spent in completing the lab exercise; (5) students' level of understanding of the targeted cloud computing principles; (6) students' application of the required skills and confidence in their abilities after finishing these labs; (7) students' development of related skills - problem solving, critical thinking, creative thinking, ability to think independently; (8) students' perceptions of their learning; (9) students' assessment of the lab as a valuable part of the course.

Evaluation Results and Analysis The evaluation was done in three semesters during 2014-2015 for two courses: CEG7380: Cloud Computing and CEG4360: Distributed Systems and Cloud Computing. The enrollments are 9, 42, and 29, respectively for CEG4360 in Spring 2014, CEG7380 in Fall 2014, and CEG4360 in Spring 2015. Figure 4 shows the distribution of the evaluated labs. We select some evaluation metrics to show the efficiency and effectiveness.

Labs	2014 Spring /CEG4360	2014 Fall / CEG7380	2015 Spring / CEG4360
AWS	X		
Hadoop/MapReduce	X	X	X
Advanced AWS		X	
Pig Programming		X	
Data Privacy		X	
Web App Resource			X

Figure 4: The evaluated labs during 2014-2015.

1. How much time in total did you spend in completing the lab exercise?
2. Your level of interest in this lab exercise. (high, average, low)
3. How challenging is this lab exercise? (high, average, low)
4. How valuable is this lab as a part of the course? (high, average, low)
5. Are the supporting materials and lectures helpful for you to finish the project? (very helpful, somewhat helpful, not helpful)
6. How confident do you feel on applying the skills learned in the lab to solve other problems? (high, average, low)

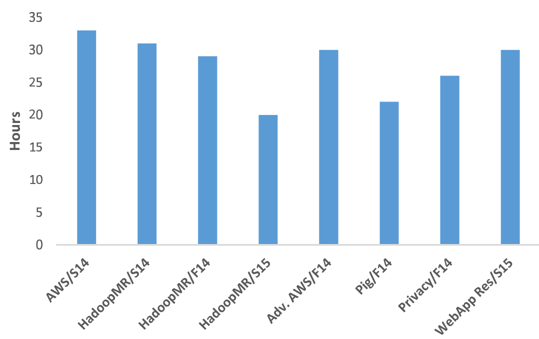

Figure 5: The average time spent on the labs by the students. Notation: Lab-Name/semester.

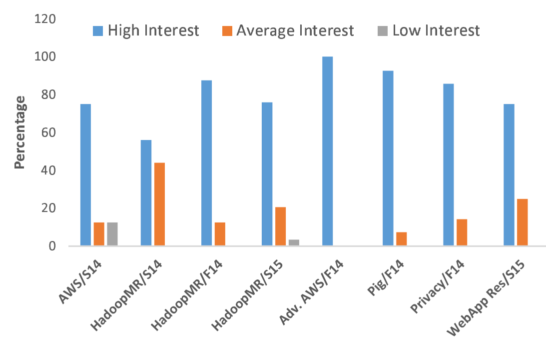

Figure 6: Students' interest on the labs.

Figure 5 shows the average time spent on each lab. These labs can be finished in about 20-30 hours. HadoopMR/S15 was revised to be easier for undergraduate students, and thus its average time is reduced to 20 hours, compared to other versions of HadoopMR. Figure 6 shows that overall students are very interested in these labs. The HadoopMR/S14 lab was a little bit too difficult for undergraduate students, which was also indicated by the difficulty evaluation as shown in Figure 7. The difficulty level of the HadoopMR labs and the WebApp Res. lab can be adjusted in the future design. Other labs have reasonable difficulty levels. Students all highly valued these labs as shown in Figure 8. Students' confidence in applying the learned skills (Figure 9) seems correlated with

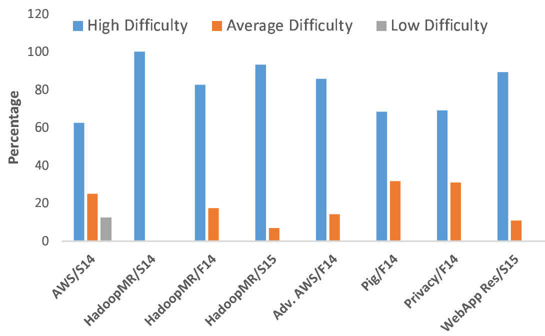

Figure 7: The lab difficulty ratings.

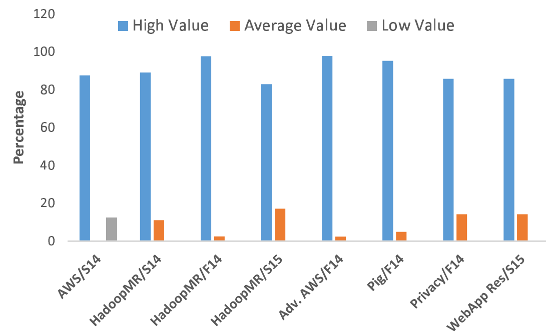

Figure 8: How valuable are these labs?

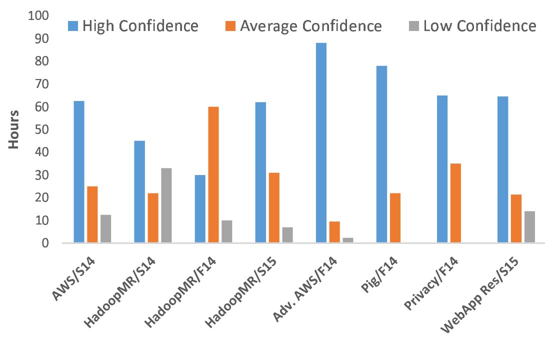

Figure 9: Students' confidence in applying the learned skills.

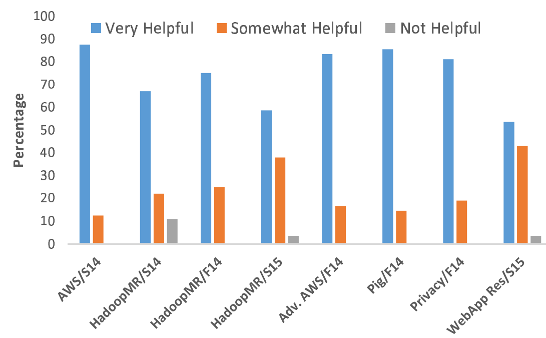

Figure 10: How helpful are the supporting materials?

their feeling on difficulty levels, which is reasonable. Finally, as Figure 10 shows, the majority of students consider the supporting materials are very useful. However, we see there are still rooms for us to improve.

6 Conclusion

With the increased popularity of cloud computing technology and applications, cloud computing education has started to follow up to meet the market demands. Cloud computing, similar to most computer science courses, has a major focus on the training of practical skills. However, only a few labs so far were developed and accessible to the public. In this paper, we report the design and evaluation of the CUTE labs for cloud computing education. The CUTE labs comprehensively cover the four aspects of cloud computing: cloud systems, big data computing, cloud economics, and cloud security and privacy. So far, we have developed 13 labs. 11 of the 13 have been deployed during 2009-2015 and 6 of the 11 have been evaluated during 2014-2015. Six evaluation metrics are used, mainly based on students' surveys. The results show that students like and benefit from these labs, but a few labs might be too difficult. Overall, we believe these labs have successfully reached their initial design goals. We will continue to improve them and develop more high-quality labs.

7 Acknowledgment

This work is supported by the National Science Foundation under Grant No. 1245847. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Rakesh Agrawal, Anastasia Ailamaki, Philip A. Bernstein, Eric A. Brewer, Michael J. Carey, Surajit Chaudhuri, AnHai Doan, Daniela Florescu, Michael J. Franklin, Hector Garcia-Molina, Johannes Gehrke, Le Gruenwald, Laura M. Haas, Alon Y. Halevy, Joseph M. Hellerstein, Yannis E. Ioannidis, Hank F. Korth, Donald Kossmann, Samuel Madden, Roger Magoulas, Beng Chin Ooi, Tim O'Reilly, Raghu Ramakrishnan, Sunita Sarawagi, Michael Stonebraker, Alexander S. Szalay, and Gerhard Weikum. The claremont report on database research. *SIGMOD Record*, 37(3):9–19, 2008.

[2] Rakesh Agrawal and Ramakrishnan Srikant. Privacy-preserving data mining. In *Proceedings of ACM SIGMOD Conference*, pages 439–450, Dallas, Texas, 2000. ACM.

[3] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz andAndy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. Above the clouds: A berkeley view of cloud computing. *Technical Report, University of Berkerley*, 2009.

[4] Ken Birman. CS5412: Cloud computing. http://www.cs.cornell.edu/courses/cs5412/2012sp/, 2011.

[5] Keke Chen. CS499/699: Cloud computing. http://www.cs.wright.edu/ keke.chen/cloud/, 2009, 2010.

[6] Keke Chen, Ramakanth Kavuluru, and Shumin Guo. Rasp: Efficient multidimensional range query on attack-resilient encrypted databases. In *ACM Conference on Data and Application Security and Privacy*, pages 249–260, 2011.

[7] Keke Chen and Ling Liu. Geometric data perturbation for outsourced data mining. *Knowledge and Information Systems*, 29(3), 2011.

[8] Keke Chen, James Powers, Shumin Guo, and Fengguang Tian. Cresp: Towards optimal resource provisioning for mapreduce computing in public clouds. *IEEE Transactions on Parallel and Distributed Systems*, 25(6), 2014.

[9] Keke Chen, Huiqi Xu, Fengguang Tian, and Shumin Guo. Cloudvista: Visual cluster exploration for extreme scale data in the cloud. In *SSDBM*, pages 332–350, 2011.

[10] Timothy Chou. CS309A: Cloud computing. https://sites.google.com/site/stanfordcs309a/, 2011.

[11] Jeffrey Dean and Sanjay Ghemawat. Mapreduce: Simplified data processing on large clusters. In *OSDI*, pages 137–150, 2004.

[12] Peter J. Denning. Great principles of computing. *Commun. ACM*, 46:15–20, November 2003.

[13] Nathan Eddy. Small businesses lack knowledge about cloud computing: Report. http://www.eweek.com/c/a/Midmarket/Small-Businesses-Lack-Knowledge-About-Cloud-Computing-Report-646801/, 2011.

[14] Richard N. Katz (Editor). *The Tower and the Cloud*. EDUCAUSE, 2008.

[15] Armando Fox. Cloud computing in education. http://inews.berkeley.edu/articles/Spring2009/cloud-computing, 2009.

[16] Tim Greene. Survey: Most businesses haven't mastered cloud security. *NetworkWorld, http://www.infoworld.com/d/cloud-computing/survey-most-businesses-havent-mastered-cloud-security-280*, 2009.

[17] Ragib Hasan. CS600.412: Security and privacy in cloud computing. http://www.cs.jhu.edu/ ragib/sp10/cs412/, 2010.

[18] Karla Hignite, Richard N. Katz, and Ronald Yanosky. Shaping the higher education cloud. http://www.educause.edu/Resources/ShapingtheHigherEducationCloud/205427, 2010.

[19] U Kang, Charalampos E. Tsourakakis, and Christos Faloutsos. Pegasus: Mining peta-scale graphs. *Knowledge and Information Systems (KAIS)*, 2010.

[20] Ron Kohavi, Neal J Rothleder, and Evangelos Simoudis. Emerging trends in business analytics. *ACM Communication*, 2002.

[21] David A Kolb. *Experiential Learning: Experience as the Source of Learning and Development*. Prentice Hall, 1983.

[22] Vivek Kundra. Tight budget? look to the 'cloud'. http://www.nytimes.com/2011/08/31/opinion/tight-budget-look-to-the-cloud.html, 2011.

[23] Jimmy Lin. INFM718G/CMSC838G: Data-intensive information processing applications. http://www.umiacs.umd.edu/ jimmylin/cloud-2010-Spring/, 2010.

[24] Tim Mather, Subra Kumaraswamy, and Shahed Latif. *Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance*. O'Reilly, 2010.

[25] Biswanath Panda, Joshua S. Herbach, Sugato Basu, and Roberto J. Bayardo. Planet: Massively parall learning of tree ensembles with mapreduce. In *Proceedings of Very Large Databases Conference (VLDB)*, 2009.

[26] John Powell. Cloud computing what is it and what does it mean for education? http://erevolution.jiscinvolve.org/wp/files/2009/07/clouds-johnpowell.pdf, 2009.

[27] J. J. Rehr, J. P. Gardner, M. Prange, L. Svec, and F. Vila. Scientific computing in the cloud. http://arxiv.org/ftp/arxiv/papers/0901/0901.0029.pdf, 2008.

[28] Michael C Schatz, Ben Langmead, and Steven L Salzberg. Cloud computing and the dna data race. *Nature Biotechnology*, 28:691–693, 2010.

[29] Reed Sheard. Cloud computing in education: A practitioner's view. http://campustechnology.com/articles/2010/09/22/cloud-computing-in-education-a-practitioners-view.aspx, 2010.

[30] Jens-S. Vockler, Gideon Juve, Ewa Deelman, Mats Rynge, and G. Bruce Berriman. Experiences using cloud computing for a scientific workflow application. In *Workshop on Scientific Cloud Computing*, 2011.

[31] Tom White. *Hadoop: The Definitive Guide*. O'Reilly Media, 2009.

[32] Matei Zaharia, Mosharaf Chowdhury, Michael J. Franklin, Scott Shenker, and Ion Stoica. Spark: Cluster computing with working sets. In *Proceedings of the 2Nd USENIX Conference on Hot Topics in Cloud Computing*, HotCloud'10, pages 10–10, Berkeley, CA, USA, 2010. USENIX Association.