MOOC and Clouds Provides an Opportunity for Teaching, Learning Environment to Students

Author1:
Ashwini.E.M
CS Department
PESIAMS, Shimoga
ashwini.em@gmail.com

Author2:
Chandrasekhar. K. L
Librarian
PESITM, Shimoga
chandrusdm@gmail.com

Author3:
Tejaswini .V. R
CS Department
PESIAMS, Shimoga
tejaswinivr @pestrust.edu.in

Abstract - A knowledge of parallel and distributed computing is important for students needing to address big data problems in later jobs in industry or academia. However, many campuses do not offer courses in these important areas due to curriculum limitations, a lack of faculty expertise, and limited instructional computing resources. MOOC and Clouds provide an opportunity to scale learning environments and to help institutions needing an advanced curriculum. In this paper, here we discuss a course offered at Indiana University and use it as a model for improving curriculum at institutions who cannot easily provide the needed courses themselves. By considering this example, we can also implement this in our institutions.

I. INTRODUCTION

Parallel and distributed computing is becoming ever more important with the exponential growth of data production in areas such as the web and Internet of Things. Further modern computers are equipped with multiple processors that need to be utilized efficiently. On the other hand, clouds are becoming the standard computing platform for running both applications as well as data analytics. With these trends it becomes increasingly important for the next generation of software engineers and researchers to be familiar with distributed and cloud computing paradigms and how they can be applied in practice and often in parallel fashion. Unlike the academy where one focuses on the fundamental computer science problems, cloud computing involves many technologies and software tools that are widely used by industry and academia for real-world applications that are now part of everyday life for billions of people. These include Internet-scale web search, e-mail, online commerce, social networks, geo-location and map services, photo sharing, automated natural language translation, document preparation and collaboration, media distribution, teleconferencing and online gaming.[14]

However the underlying fundamentals of these techniques are coming from different computer science disciplines including distributed and parallel computing, databases and computer systems architecture. A well-rounded course of cloud computing should cover each of these areas and explain them in the context of cloud computing. To gain practical experience on cloud computing, a student has to master many different technologies that are based on these principles. In order to facilitate such a learning environment, Indiana University developed the Cloud Computing online course 1. This course has been taught by faculty for several years both in-class for residential students and for online students. The course is offered as part of the curriculum for Computer Science graduate program at Indiana University and students from the Data Science graduate program. Intelligent System Engineering and Library Science are also given the opportunity to take the course. The population of online students is geographically located world-wide, from London, France, Germany, India to Indianapolis. Most of the students are professionals who take online classes to update their knowledge and skills or earn a degree.

A primary goal of the course is to maintain the same standard as the residential course for the online course. Since this is a programming intensive systems course, it is especially challenging due to limitations on the face to face interactions with the online students, diverse background of students and the deep technical knowledge required by the course. The students are expected to have general programming experience with Linux and ability in the Java (2-3 years) programming language and scripting. A background in parallel and cluster computing is considered a plus but not required. The statistics present in this paper are related to the latest version of the online course which saw the largest attendance so far with about 160 students, where 100 were residential students and the rest were online students. The popularity of the cloud computing topic follows from major corporations including Microsoft, Amazon, Google, IBM, Facebook and Twitter, which provide infrastructure, tools or applications in Clouds. Business, government, academia and individuals use public or private cloud-based solutions for storage and applications. The course has been taken as a model by other institutions to introduce cloud computing to their own students. This is facilitated by the availability of online course materials.[14]

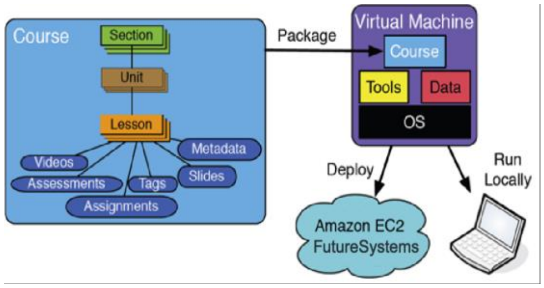

Fig. 1. Model for the MOOC Class Content and Delivery using Cloud

II. COURSE PROGRESS AND COURSE ASSOCIATION

The classes are aimed at teaching the basic principles of parallel and distributed computing and explore the application of these in practice in cloud environments. This is a graduate level course with large emphasis on programming and expects prior knowledge of programming in order to be successful. The course follows the cloud computing text book [1]. By the end of this course, students are expected to learn key concepts in cloud computing and have enough hands-on programming to be able to solve data analysis problems on their own. The organization of the course is shown in Fig. 1.

A. Course Content - The class uses the Google Course builder as the content hosting platform. Google Course builder provides an easy way to host course content. Its source code is distributed under the Apache License version 2 and is free to modify and redistribute.

The class's content is mainly lecture videos hosted in YouTube. Text version of the content is also possible. The class has been structured as a set of units. Each unit contains a set of lessons. Each lesson is a video plus some text description. Each lesson can be followed by a simple activity. The lecturer creates an activity as a JavaScript file. The activity contains simple multiple-choice questions or text based answer questions with specific answers. Between units there can be course assessments. These assessments can be quizzes, midterm exam and final exam. They also have the same format as activities followed by lessons and feature multiple choice questions and simple text based answer questions. The activities and assessments can be graded and the scores are displayed in the student profile.

The course consists of six units starting with cloud computing fundamentals and then move on to infrastructure as a service (IaaS), Platform as a service (PaaS) and cloud data storage and Internet of Things applications. Each unit consists of multiple lectures with videos. The videos were recorded by the lecturer with the help of a professional staff for video recording and editing. It took a lot of effort and time to get the videos properly recorded in the first time of offering the course. After the initial videos were finalized it was relatively easy to add more content or update the videos for later offerings of the course.

B. Projects - The course was offered with a comprehensive set of cloud application projects that are interlinked together. The overall goal is to build a web search engine from scratch. Students can use various tools and build the system one component at a time using cloud based data analytic platforms.

The projects start with a small activity that involves configuring Hadoop and running a simple Hadoop program. The first building block of the search engine expects students to write a pagerank [6] algorithm in Hadoop to assign an importance to web pages. Next the HBase distributed storage is introduced to the students and the course expects them to write a program to load the data into HBase as well as create an inverted index from word to page to facilitate the search. The next step is to combine the results from pagerank and use the inverted index to do actual searches. Apart from these projects the students are expected to implement a standard machine learning algorithm using the Harp [7] machine learning platform.

C. Assignments & Exams - Assignments were mainly focused on testing the basic knowledge about the subject matter. Most questions are selected from the text book. Assignments were given weekly or twice a week and had a turnaround time of week.

D. Student Evaluations - Students are evaluated based on their performance on programming projects, written assignments and exams. The exams are focused on the core concepts of cloud computing and related underlying principles. For online students, the exams are conducted using the Canvas platform and the Adobe connect video conferencing.

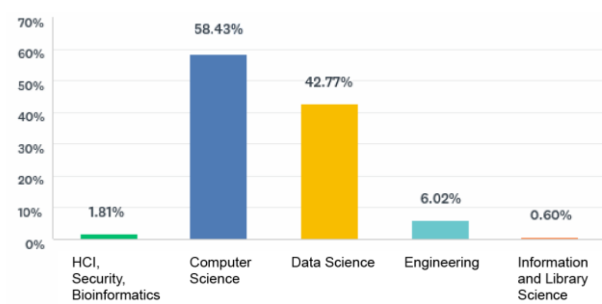

Fig. 2. Departments where the course is cross-listed among five different programs: Informatics, Computer Science, Data Science, Intelligent Systems Engineering, Information and Library Science.

III. COURSE SCALING AND TECHNIQUES

This course needs to explore several Hadoop oriented technologies in dealing with Big Data on Cloud Computing.

Although prior knowledge of the field is desirable, most students expressed their lack of experiences on these new technologies as they are in the first year of graduate study. We also observed that students eager to learn a wide range of knowledge and experiences about parallel computing with particular software such as Apache Hive, Spark, Pig and Lucene being of interest.

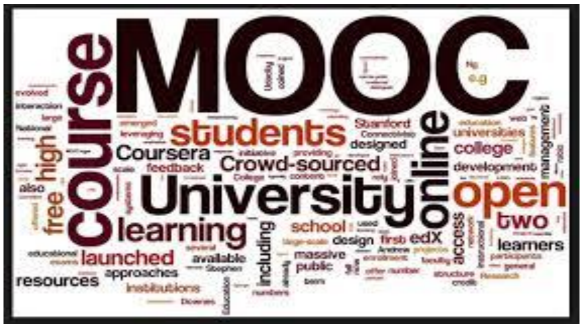

Fig. 3. Different ways of MOOCs benefits to teachers

A. Forums : Since the course is offered for a large number of online and residential students from different time zones and different professions, providing interactive support of course materials, especially about hands-on projects with code implementations is one of the challenging tasks for lecturer. Considering with several options for class forums which is a vital part of the course. Because the large class size, an lecturer is not always possible to solve problems encountered by individual student in person. In previous years the course was run with Google Forums2, Indiana University internal forums and Piazza3 forums and we found Piazza to be the best option. The web-based tool Piazza is mainly used for the communication between lecturer and students and among students and our statistics indicate that 84% of questions received responses within 61minutes in average.

B. Hands-on Labs: The course is organized with twice a week projects to encourage active developments in source code writing and connect between an literature in a textbook and the latest technologies. The fundamental pedagogy underlying these hands-on projects is to embrace new experiences in learning both theory and practice with minimal barriers, for example, learning a new programming language or preparing computing environments with recent software tools, which takes effort and time to achieve.

Many students have at least 1 or 2 years language experiences among Java, C#, C++, C and Python which are abundant to start basic code developments in most assignments. One of the challenging activities in teaching from previous classes is building a controlled experimental environment over different computing platforms. We built a virtual machine image to avoid such difficulties and the choice of computing environments is given to students based on their confidence level. The VM image is able to run on the student's desktop via VirtualBox. The transition from using a desktop to run jobs on the cloud environment is a steep learning curve. The labs provide students with step-by-step instructions on how to install and configure the Hadoop.

C. Online Meetings: To assist the questions from students regarding both class content as well as projects, online meet ups were conducted every week. These were one hour sessions mainly steered by associate lecturer. In early course offerings, with only a small number of students the Google Hangouts platform was the choice for online meetings. But with larger classes we switched to Adobe Connect platform, provided by Indiana University for online courses. Every such online meeting is recorded and available through YouTube for later viewing by the students. We find that such videos are helpful in the subsequent runs of the course as well. Video conferencing tool Adobe Connect (now replaced by Zoom) is also provided for weekly class lab session and office hours to instruct how to complete course assignments with step-by-step tutorials and provide individual feedback. Recordings for these sessions have been made for self study in case that students need to re-visit materials covered in those sessions.

D. Content Repository: One innovation of this project is to build on our extensive experience with online education and its technologies to use MOOCs technologies and build an open source community X-MOOC repository to explore a modular and customizable process for storing, managing, and sharing course content and learning materials. The developer of the course found the need to share content among different courses run by different instructors. In order to do so, a MOOC platform should be able to share course content among different courses. As part of this course, we have developed technology on top of edX MOOC platform to share content among different courses and we have moved the content of this course which is currently in Google Course Builder to the edX platform. This development will allow instructors from different universities to easily share course content and quickly create new courses that modify the old one. Link function adds a link to related course videos in assessments. This function can navigate students through specific course content for review. It is particularly helpful when students make mistakes in a quiz or an exam but are unclear about the missing knowledge. The instructors can provide a list of keywords and their associated video tags for each question. If a student chooses an incorrect answer, these video links will automatically appear as hints (buttons) under the question.

IV. COURSE OUTCOMES AND EVALUATIONS

In addition to standard institutional class evaluations, we conducted a post course survey to gather feedback on the details of course content and measure the growth of students in obtaining the knowledge and skills. For the cloud computing course we wanted to know the preference of students in using a VM instead of a distributed environment for the projects. The fact that VM is easier to setup and run programs rather than using a distributed cluster.

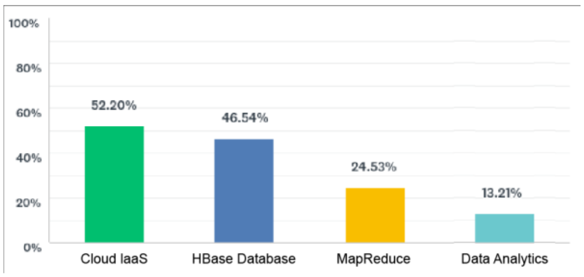

Fig. 4. Prior Knowledge at the first day of class to learn during the course.

VI. CONCLUSION AND FUTURE WORK

The Cloud Computing course has been offered for many years to the residential computer science graduate students at the Indiana University and has observed high enrollments each time. By considering this real time example of Indiana University, in India we can also step forward to use Clouds and MOOCs benefit and also give benefit to our students also. The course offers a mix of core concepts of distributed and parallel computing along with their practical applicability. This combination has been proven to be successful in teaching a diverse group of students who are primarily looking towards industry which increasingly demands engineers with experience in distributed and parallel computing domains. Faculty and IU support have helped develop a curriculum for remote sensing materials and this will allow other institutions within the AMDI community. Clouds and online MOOCs offer cutting-edge technologies to enhance traditional computational science curriculum and research with next-generation learning descriptions.

For future work, In India all our universities and colleges will continue modernizing curricula that is suitable for our next generation workforce development and connect to the community by systematically introducing multiple courses, teacher training, research support and electronic resources across the ADMI MSI and other teaching university networks.

REFERENCES

[1] K. Hwang, J. Dongarra, and G. C. Fox, Distributed and cloud computing: from parallel processing to the internet of things. Morgan Kaufmann, 2013.
[2] T. White, Hadoop: The Definitive Guide, 1st ed. Sebastopol, CA, USA: O'Reilly Media, Inc., 2009.
[3] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST), May 2010, pp. 1–10.
[4] L. George, HBase: the definitive guide: random access to your planetsize data. " O'Reilly Media, Inc.", 2011.
[5] J. Callan, M. Hoy, C. Yoo, and L. Zhao, "Clueweb09 data set," 2009.
[6] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.
[7] B. Zhang, Y. Ruan, and J. Qiu, "harp: Collective communication on hadoop," in 2015 IEEE International Conference on Cloud Engineering.
[8] J. Van Orshoven, R. Wawer, and K. Duytschaever, "Effectiveness of a train-the-trainer initiative dealing with free and open source software for geomatics," in Proceedings (J.-H. Haunert, B. Kieler and J. Milde, Eds.) of the 12th AGILE International Conference on Geographic information Science, 2009.
[9] B. Fishman, S. Best, J. Foster, and R. Marx, "Fostering teacher learning in systemic reform: a design proposal for developing professional development." 2000.
[10] J. H. van Driel, D. Beijaard, and N. Verloop, "Professional development and reform in science education: The role of teachers' practical knowledge," Journal of Research in Science Teaching, vol. 38, no. 2, pp. 137–158, 2001. [Online]. Available: http://dx.doi.org/10.1002/1098-2736(200102)38:2¡137::AID-TEA1001¿3.0.CO;2-U
[11] D. Hestenes, "Toward a modeling theory of physics instruction," American Journal of Physics, vol. 55, no. 5, pp. 440–454, 1987. [Online]. Available: http://dx.doi.org/10.1119/1.15129
[12] H. Borko, "Professional development and teacher learning: Mapping the terrain," Educational Researcher, vol. 33, no. 8, pp. 3–15, 2004. [Online]. Available: http://dx.doi.org/10.3102/0013189X033008003
[13] X-mooc repository: Curriculum enhancements with cloud and mooc for online learning. [Online]. Available: http://cloudmooc2.soic.indiana.edu/

[14] Teaching, Learning and Collaborating through Cloud Teaching, Learning and Collaborating through Cloud Computing Online Classes by
Judy Qiu_, Supun Kamburugamuve_, Hyungro Lee_, Jerome Mitchell_, Rebecca Caldwelly, Gina Bullockz, Linda Haydenx _School of Informatics, Computing, and Engineering, Indiana University Bloomigntonfxqiu, skamburu, lee212, jeromitcg@indiana.edu
yWinston-Salem State Universitycaldwellr@wssu.eduzNorth Carolina Agricultural and Technical State University
glbulloc@ncat.edu xElizabeth City State University
haydenl@mindspring.com