On the Integration of Big Data and Cloud Computing Topics.[‡]

Debzani Deb, Dept. of Computer Science, Winston-Salem State University, debd@wssu.edu

ABSTRACT

With the fast growth of big data and cloud computing paradigm, we argue that each and every CS and IT students should be equipped with foundation knowledge in this collective paradigm and should possess hand-on-experiences in managing big data applications in clouds to acquire skills that are necessary to meet current and future industry demands. This poster presents our research that proposes gradual and systematic integration of big data and cloud computing related topics into multiple core (required) courses of CS/IT curriculum. The NSF funded study will be useful for CS/IT students and their instructors as it will identify big data and cloud computing related topics that are important to cover, find a sequence of the prescribed topics that can be incorporated into existing core courses most effectively, and suggest specific core courses in which their coverage might find an appropriate context. The poster will further identify the major challenges this intervention may encounter and will provide a deeper analysis of them. Finally, the poster will describe our experience of implementing one such course with proposed interventions during Fall of 2016 semester. The pre- post- test results that measure student opinion and understanding about big data and cloud computing topics will be presented in the poster which demonstrates improved student interest and learning.

SIGNIFICANCE AND RELEVANCE OF THE TOPIC

Big data and cloud computing (BDCloud) collectively offer a paradigm shift in the way businesses are now acquiring, using and managing information technology. With the fast growth of this paradigm, businesses are struggling to find experienced people who not only have the deep analytical skills, but also have the data hosting, storage and management skills to effectively leverage this collective model. This observation is supported by the recent prediction made by the International Data Corporation (IDC), where the forecast is that by 2020 big data staffing shortages will expand from analysts and data scientists to include architects and experts in data management. With an estimated number of 50 billion devices that will be networked by 2020, Internet of Things (IoT) will only alleviate this demand as it will be one of the main sources of big data, and cloud will be an enabler for storing it for long time and for performing complex analysis on it.

We believe that CS and IT students should be equipped with foundation knowledge in this collective paradigm and should possess hand-on-experiences in managing big data applications in clouds to acquire skills that are necessary to meet current and future industry demands. Academia, although responded slowly, recently had taken multiple approaches to fill this critical void and to produce a satisfactory number of well-trained professional. Our extensive literature survey identifies three such approaches from Academia. Firstly, a large number of initiatives focus on offering a specialized "Data Science" course to CS/IT majors, where students are primarily taught data acquisition, cleaning, analytical and visualization skills. While this particular approach helps students developing skills related to transforming data into knowledge, it does not provide them with concepts and experiences related to hosting, storing, designing and deploying applications within performance and budgetary constraints. A second category of approaches integrates big data and/or cloud computing topics in existing CS/IT courses (mostly non-required) such as Parallel and Distributed computing, High Performance Computing, Networking, Cybersecurity etc. These interventions are often sporadic, mostly ignited by instructor's own interests and experiences on BDCloud topics and therefore were not capable of garnering a substantial student interest and knowledge on this collective paradigm. Finally, a number of research-intensive universities offer specialized standalone courses such as "Cloud computing", "Big data management" etc., where the abovementioned collective paradigm is addressed in a greater extent. However being a special topic course and offered at a handful of universities, only a small number of enrolled students perceive the benefit.

[‡] Is supported by NSF award #1600864

We argue that for substantial coverage of big data and cloud computing concepts and skills, the interventions need to be gradual and systematic and should be integrated into multiple core (required) courses of CS/IT curriculum. This research aims to identify a sequence of core CS/IT courses, where BDCloud related can be integrated with the goal of providing an increasing and systematic depth of coverage throughout the curricula. This NSF funded project will support a multi-course intervention during the next three years and will continuously assess student interest and learning. The study results will be useful for CS/IT students and their instructors as it will identify BDCloud topics that are important to cover, find a sequence of the prescribed topics that can be incorporated into existing core courses most effectively, and suggest specific core courses in which their coverage might find an appropriate context.

This study also aims to identify pedagogical and other challenges related to the proposed gradual approach followed by possible remedies for them. As an example, one can easily argue that CS/IT core courses are already exhausted with all guidelines and requirements imposed by ACM, ABET, CS departments etc. and instructors typically have no or very little room for adding more topics. However, the integration can be justified by the recent 2013 ACM/IEEE Curriculum guidelines which introduced Parallel and Distributed Computing as a first class knowledge area for the first time and the concepts of BDCloud is well suited to cover under this area. One can further argue that teaching students about big data and cloud computing systems typically requires resources far beyond of that available to an individual course or department. While this is a serious challenge, there are many possible solutions available for the instructors to pursue such as utilizing a virtual cluster hosted on a physical machine, or using NSF supported free Cloud infrastructure such as Chameleon and CloudLab, or exploding low-cost Amazon web services (AWS) in Education etc. As we progress with this research, we hope to explore more such challenges that an instructor may encounter while adopting BDcloud topics into existing curricula.

The study will be adopted to CS and IT courses in Winston-Salem State University (WSSU), a HBCU, that serves unique group of students as 71% of its student population are female and 72% are African American. This research will restructure few core courses in order to enhance the educational experiences of CS and IT majors. These activities are expected to enrich students with necessary knowledge, tools and hand-on-experiences on big data and cloud computing so that they are better equipped to address challenges in computational and data-enabled science that nation is currently seriously lacking. In order to intrigue interests among this unique group of students, we will exercise innovative pedagogy such as active learning, problem-based, and case-study based learning while introducing the BDCloud topics. Our goal of this curriculum enhancement is to create a well-rounded and professional pool of minority students who are better prepared to handle advanced degrees and careers in STEM and who can take on multidisciplinary global challenges.

CONTENT

The poster will present our experience and insight gained so far in this research and will showcase the proposed curriculum encashment. More specifically, the poster will present the important BDCloud related topics and the breakdown of these topics among four core CS/IT courses. The poster will further identify the major challenges this intervention may encounter and will provide a deeper analysis of them. Lastly, we will describe our experience of executing one such course with proposed interventions during Fall of 2016 semester. The pre- post-test results that measure student opinion and understanding about BDCloud topics will be presented in the poster which demonstrates improved student interest and learning.