

An Alternative Approach to Teaching Bigdata and Cloud Computing Topics at CS Undergraduate Level

Debzani Deb, Muztaba Fuad, Keith Irwin, Department of Computer Science

Winston-Salem State University

debd@wssu.edu, fuadmo@wssu.edu, irwinke@wssu.edu



WINSTON-SALEM
STATE UNIVERSITY

Rationale and Proposition

- **Big Data** and **Cloud Computing** collectively offer a paradigm shift in the way businesses are now acquiring, using and managing information technology.
- Every CS/IT students should be equipped with foundation knowledge in this collective paradigm and should possess some hands-on-experiences in developing and managing big data applications in the cloud platform.
- **We argue that for substantial coverage of big data and cloud computing concepts and skills, the relevant topics need to be integrated into multiple core courses of undergraduate CS curriculum rather than creating additional standalone core or elective courses.**
- Our approach to including these topics is to develop **learning modules** for specific core courses in which their coverage might find an appropriate context.

Learning Modules

- Two 75-minute classes (one week) were devoted to cover each module in respective courses.
- Each module has specific learning goals, lessons plans, and assessment resources.
- Three such modules are developed and deployed in **Computer Architecture, Analysis of Algorithm** and **Database Management** class.
- Interested readers are advised to explore resources posted at the **project website**.

Acknowledgment



NSF RIA Award #1600864



Feedback @ <http://ibigcloud.altl.org/>

“Computer Architecture” Module

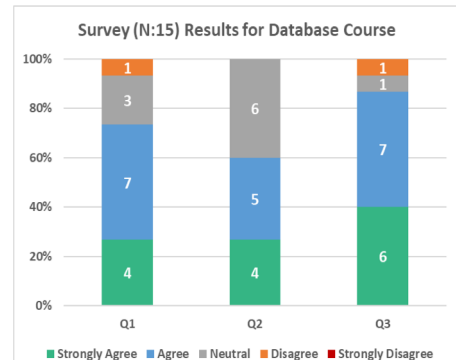
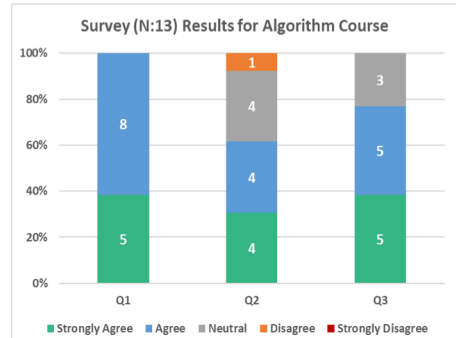
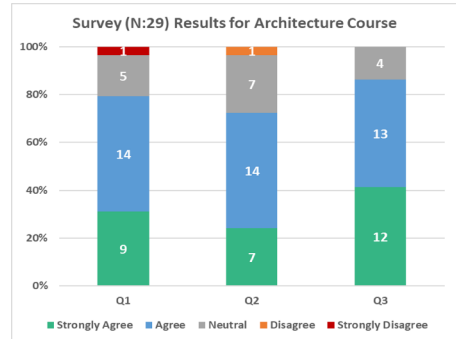
- **Learning Goal:** Develop basic understanding with virtualization and cloud computing frameworks and gain hands-on experience in using AWS cloud.
- The project involves exploring Amazon EC2 web services by creating various EC2 instances and by benchmarking these instances to determine how performance scales with the different virtual machine types. Students are provided with various benchmarking applications (CPU, IO) and are required to represent their results graphically and to report on them critically.
- Students were able to accomplish the project with 40% of them achieving A grade, 20% achieving B grade, 15% securing C grade, and 25% attaining D/F grades.

“Analysis of Algorithm” Module

- **Learning Goal:** Develop basic experience with big-data analysis on cloud platforms with performance and cost constraints using Spark/Hadoop platforms.
- The project involves modifying and developing MapReduce applications, and deploying Spark applications in the cloud while focusing on resource provisioning and performance vs. cost tradeoff.
- All students were able to modify MapReduce code to meet new requirements, whereas 85% of them were able to develop MapReduce program from scratch, and only 30% of them successfully completed the cloud deployment and analyzing cost vs. performance task.

“Database Management” Module

- **Learning Goal:** Recognize that the use of SQL has been significantly extended with the advent of Big data systems and develop hands-on experience in using Spark SQL within these systems.
- The project involves developing Spark application that analyzes large data sets and use Spark SQL to query the data. In this project, students explored situations where the schema could be inferred from the data as well as the situation where it must be enforced programmatically.
- Students were able to accomplish Task 1 (infer schema automatically) while 63% of them attained A grade, 25% of them attained B grade and 12% attained C grade. Task 2 (define schema programmatically), however was successfully completed by only 12% of the students.



Q1 – I found topic X interesting.

Q2 – If a friend asks me what topic X is, I will be able to explain for 2-3 minutes.

Q3 – I would like to learn more about topic X.