

BigData Platform - Syllabus

Section 001, 3 Credits

Wednesday, 09:00-12:00



Instructor

Dr. Preethi Ananthachari

Email

preethi.ga@wsu.ac.kr

Office Hours & Location

W19 - #230 – By Appointment

General Information

Required Text: Hadoop: The Definitive Guide, by Tom White, Released June 2009

Publisher(s): O'Reilly Media, Inc, ISBN: 9780596521974

Description:

This course covers (1) the principal features of the Hadoop tools (2) using mapreduce program to collect, process, and analyze data, and (3) Yarn Job scheduler, Hive and HBase. Hadoop is an open-source framework that allows to store and process big data in a distributed environment across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. This course also provides opportunities to develop an understanding of how programming is situated in and reflects broader social structures, constructs and issues, e.g. race, class or gender. Programming is often viewed as a value-neutral technical skill. Though, the communal and traditional impacts of information and technology are central concepts in our field, and the growing awareness of issues like algorithmic bias, ethical/unethical uses of algorithms and disparities in opportunities in tech jobs require that any informed professional needs to understand the larger context of programming. This is important to be ethical professionals and to be successful in the workplace. Through readings, discussion and writing, we will critically examine issues of racism, sexism and other forms of power and oppression that are pervasive in programming and related technical activities, and discuss what companies and individuals are doing to improve programming practices and professional work environments.

Teaching Methods

*This course assumes a basic understanding of procedural programming, and begins with a comprehensive review of Bigdata fundamentals—including data types, variables, loops, and conditionals—that is designed to deepen your mastery of these concepts. Alternatively, if you have worked with a language such as Java, , you should be able to apply that knowledge to learning Hadoop. The later parts of the course will cover certain topics in program design and programming best practices (documentation, testing, etc.) that are a necessary part of producing complex, reliable, and maintainable applications. Lecture notes will be uploaded via online. Students should expect that critical information will be posted on the Smart LMS or sent through messages from Smart LMS. You are expected to check Smart LMS frequently. (<http://smart.wsu.ac.kr>). Because of the heavy reliance on online materials, all students are required to have a computer to use. **Students need to use Computer kit.***

2. Computer: (provided by school) All software in the class is multi-platform, so Windows, Mac, and Linux systems may be accommodated. However, students will need to be able to install software onto their computers. Therefore, more limited devices like Chromebooks and tablets (e.g., iPads, Fire) may not be acceptable.

Prerequisites/presumed foundational knowledge and skills

There are no elective prerequisites for this course.

Core Competencies

Core Competency	Sub-Competency	Sub-Competency
Moral and ethic	Ethical Reasoning	
Social Interaction	Teamwork	
Creative and Logical Thinking	Creative Thinking	
Analytic Skill	Critical Thinking	Information Literacy
Problem Solving	Inquiry and Analysis	Problem Solving
Communication	Oral Communication	Multimedia Communication
Global Perspective	Global Learning	

Course Goals and Learning Outcomes

Upon completion of this course, students will be able to:

In this course, students will:

Explore the broad scope of Bigdata analysis.

2. Build the foundation of commodity hardware.

3. Learn mapreduce for distributed processing.

4. Acquire wordcount program skills to solve problems.

5. Develop bigdata programming techniques.

6. Apply software skills to **manipulate databases using Hive query Language etc.,**

7. Perform innovative problem solving in a project group setting.

Communicate technical facts through verbal/written methods. physical demos. and multimedia

Course Requirements & Grading

Component	Weight
Attendance	20%
Midterm Examination	20%
Final Examination	20%
Projects	25%
Participation, Lab	15%

Evaluations

Midterm Examination (In-class) (20%): The examination will be an in-class exam covering topics discussed in the class up to this point.

Final Examination (In-class) (40%): The examination will be an in-class exam covering topics discussed in the class up to this point.

Tests/Quizzes/Activities (QZ) (20%)

Final Project: One Individual and one group based project. Each student/group will write Python programs that will apply the foundational principles of computational or quantitative thinking to solve of a problem.

First project will be submitted on Week 9 - 10%

Second project will be submitted on Week 14 - 10%

Course Participation

Student participation will be required for in-class exercises and projects. Lack of participation will affect students' final grade. Course participation will be graded by 3 random quiz (5 points each) and 5 points by professor's observation.

Grading Curve

Excluding failing students, the maximum percentages of total class grades are:

A+ ~ A0: 30%

B+ ~ D0: 70%

Course Materials

Course Text:

Hadoop: The Definitive Guide, by Tom White, Released June 2009

Publisher(s): O'Reilly Media, Inc, ISBN: 9780596521974

Hadoop in Action, Second Edition, Chuck P. Lam, Mark W. Davis, and Ajit Gaddam, **MEAP**
September 2014 , ISBN 9781617291227

Additional Materials/Resources

Students are expected to bring required texts and note-taking supplies (notebook and pen, laptop computer, etc).

Course Website and Other Classroom Management Tools

Students can follow the lecture notes and all other course related issues on SMART LMS (<http://smart.wsu.ac.kr>)

Course Expectations & Guidelines

1. Academic Integrity

Students are expected to produce their own work and, when using the work of others, include clear citations. Violations of the Academic Integrity will not be treated lightly, and disciplinary action will be taken should such violations occur.

2. Student Special Needs

If a student has a condition that impacts their classroom learning (such as a hearing disability), they should discuss special arrangements with the professor. A note from the departmental office may be required. If you experience difficulty in this course for any reason, please do not hesitate to consult with me. In addition to the resources of the department, a wide range of services is available to support you in your efforts to meet the course requirements.

3. Attendance and/or Participation

It is the student's responsibility to inform their department in the event of excused absences. The professor does not handle this process. Class attendance is taken regularly and punctuality is expected.

Since major activities will start at the beginning of the course hour, you must be on time to join the activity. No late participation is appreciated. Missing 30% of classes will result in an automatic failure.

4. Extensions, Late Assignments, & Re-Scheduled/Missed Exams

The professor do accept late submission of work, operating on a “better late than never, but better never late” policy. However, expect significant deductions based on how late the work is. The professor reserve the right to deduct however much he feels is appropriate in relation to the quality of the specific assignment and your overall performance during the semester. This can be as much as little as 5%-10% or as high as 50-75%.

5. Student Use of Mobile Devices in the Classroom

IT devices/mobiles will be allowed in the classroom for online quizzes and practical activities. Also when professor will allow mobile phones in class for researching information, double-checking content, or referencing online documents. If this privilege is abused, professor will ban mobile device use all together. Failure to comply with this policy will result in the deduction of participation points. If you would like me to consider an exception to this policy, please see me individually to discuss.

6. Additional Course Policies

No Sleeping during the lectures is welcomed. That will affect your overall grade.

Additional Syllabus Components

Students are expected to bring required textbook and note-taking supplies (notebook and pen, laptop computer, etc...).

Course Schedule

Week-Date	Topics & Textbook Readings	Class Activities and Details
1 –	Introduction to BigData	<ul style="list-style-type: none">• BigData – an Overview
2 –	Hadoop Basics.	<ul style="list-style-type: none">• Virtual Machine introduction and installment.
3-	Introduction to Ubuntu,	<ul style="list-style-type: none">• Installation of Ubuntu in Virtual machine.• Linux Comman ds in Hadoop and its Usage.
4 -	Hadoop	<ul style="list-style-type: none">• HDFS, Name node, Datanode Activity: Environmental variables of Hadoop
5 –	Commodity Hardware	<ul style="list-style-type: none">• FSImage, Edit logs in Name node• Job schedule in YARN• Mapreduce Introduction
6 –	Mapreduce	<ul style="list-style-type: none">• Mapper phase• Reducer phase• Combiner phase Activity: A simple file creation in name node

7 –	Revision for Mid-term	Practice Programs for Mid-term
8 –	Midterm Exam	In Class Exam
9 –	Wordcount / Pig	<ul style="list-style-type: none"> Wordcount program using Mapper and reducer in mapreduce.
10 –	Hive Introduction / Pig Introduction	<ul style="list-style-type: none"> Hive basics and installation Pig Latin
Week-Date	Topics & Textbook Readings	Class Activities and Details
11 –	Hive query language (HQL)	<ul style="list-style-type: none"> Create table Delete table using hive commands LOAD command in Pig.
12 –	Partitioning in Hive	<ul style="list-style-type: none"> How to partition a table in hive using its columns CROSS, DISTINCT commands in Pig
13 –	Hive Buckets	<ul style="list-style-type: none"> How to create buckets in Hive and what are all the usages? Hive views and indexes Built in operators in Hive Pig Commands
14 –	Revision	Programs practice and Projects Presentations.
15 –	Final Exam / Project	In Class Exam / Project Presentations

The instructor reserves the right to modify this schedule based on new information, extenuating circumstance, or student performance. Students will be informed before any changes and due consideration will be given to grades and participation.