CS 724 - High Performance Computing and Big Data: Syllabus

Course Description
With the advent of massively parallel technologies such as NVIDIA GPU, and big data frameworks like Hadoop and Spark, it is becoming feasible to process large amount of data in a reasonable time. Parallel and distributed computing which in the past was mostly available to scientific applications is now available to other disciplines such as financial analytics, bioinformatics, health analytics, and large scale data mining.

A brief list of topics to be covered in this course is given below:

- Overview of high performance computing and big data
- NVIDIA GPU and programming model
- Hadoop framework and programming model
- Apache Spark and programming model
- Designing big data applications in different domains such as financial, social networking, and large graph analytics.

Prerequisites: C/C++, Java, and familiarity with Python, Unix, Computer architecture, Basics of Probability & Statistics, and Calculus. If you are not sure about the requirements, please contact the instructor.

Computer Requirement: Students are expected to have access to a computer running Windows OS preferably with 8 GB of memory.

Readings
The main text for this course is:


Please visit http://www.mmds.org/ for details on the text and how you can get access to an online version of the text.

As this course is covering new and very fast changing technologies, it also relies heavily on the material available on the Internet. The pointers for reading material on the Internet is listed under each module.

Course Goals and Objectives
This course objective is to introduce parallel and distributed programming principles to tackle problems dealing with large amounts of data. The course will give practical experience of using big data platforms and also how to develop applications on these platforms. The course consists of a number of programming lab exercises and homework assignments that will help students to learn how to use big data technologies. The course project is a major component of this course, which will enable students to apply their knowledge acquired in the course to develop and implement a big data application.

How the Course Works
Course Structure, Teaching and Learning Methods
Structure:
The course is divided into two units with each unit having several modules. Unit-I is the core unit that covers three parallel and distributed computing technologies for high performance computing with big data. These are: GPUs (CUDA), Hadoop, and Spark. Unit-II covers several applications of large scale data analytics on three technologies covered in Unit-I. This is a hands-on course with a number of programming exercises, home assignment, and a final project. Participation in discussion forum is mandatory and will also be graded. See discussion participation section for more details. The course project is a major component of this course, which will enable students to apply their knowledge acquired in the course to develop and implement a big data application on one or more of the technologies covered in Unit-I.

Duration:
Unit-I is covered in 6 weeks, Unit-II and project are covered in 9 weeks. The course project overlaps with Unit-II and is expected to be completed in parallel. See the schedule section for details.

Module Activity:
- Watching module overview video from the instructor
- Reading instruction material
- Completing lab exercises and submitting all required work
- Completing homework assignment (Note that not all modules have homework assignments)
- Participation in discussion forum. This is mandatory and will be graded, see discussion participation section for details

Online meeting with instructor:
- Meeting with instructor to discuss project at various stages of completion (see Project section for details)
- Presenting the project along with demo (see Project section for details)

Grading Criteria
Your grade will be based on a total of 100 points with the following distribution:

- Homework assignments: 30 points (5 assignments)
- Lab Exercises: 10 points (9 assignments)
- Discussion Participation: 10 points (12 modules post/respond)
- Project: 50 points

Project
The course project is a major component of this course, which will enable students to apply their knowledge acquired in the course to develop and implement a big data application. Several sample projects that were judged to be of superior quality from previous years will be available via Blackboard. Project completion will occur in several phases.

Project Kick Start Phase (During 8th week): Schedule a one-on-one meeting with the instructor to discuss the initial project idea. Please watch out for a doodle invitation in your email during that week for scheduling an appointment.

Project Finalize Phase (During 10th week): You need to submit your project description. The project description should have the following structure:
Project Title:

Project Objective (three or four lines):

Project Description: One paragraph describing the large data application on a big data framework. You should clearly establish that the application will benefit using a parallel and distributed technology.

Please submit a plain text file with the above structure. You need to schedule a meeting to get feedback on the project. Please watch out for a doodle invitation in your email during that week for scheduling the meeting. Based on the feedback, you will need to revise and submit your project description.

**Project Design Submission Phase (During 12th week):** The project design should give implementation details along with a high-level description of the algorithm you will implement. For a high-level description you can use pseudocode or flow charts. The design should clearly identify how you plan to test the application and what data set you will be using for this testing. If you have any questions or concern you can post your question to the instructor in the forum titled: “Problems and Concerns”.

**Project Demo (During exam week):** I will be allocating one or two days for the project demo. Please watch out for a doodle invitation in your email during that week for scheduling the demo. The project demo consists of two parts: presentation, and demonstration of working of the project. The presentation should include an overview of the big data application, approach taken, implementation details, experimental results, and lessons learned.

**Discussion Participation**

**Discussion Board**
This is the place where course participants consult one another, discuss topics and certain assignments. Participants post documents and others respond to them, resulting in a discussion thread. The participation in the discussion board is mandatory and accounts for 10% of your grade. Note that postings are organized in different forums. There is a forum for every module. If you want to post a question or an assignment related to a module, you need to post in the corresponding forum. Besides these standard forums, there is one special forum: Problems and Concerns. In this forum, you will post any general problems and concerns. Your instructor will periodically scan these postings and respond to them. Note that questions related to specific module content should be posted in the discussion forum related to that topic. Discussion points will be accumulated for every module except theintroductory week (Module 1) and the "Project Kickstart" (Module 6) modules; a discussion board is provided for student use but these two will not be graded.

**Discussion Participation Points**
The discussion group is a critical component for this course. To encourage participation in the discussion forum, I have assigned Discussion Participation Points (DPP) based on your level of participation.

Posting a question in the discussion group -- 4 DPPs (out of 10)
Posting a comment or a solution to an existing question -- 6 DPPs (out of 10)

Your DPPs will not be based on the correctness of the posted solution, any honest attempt will be rewarded. However, points will not be awarded for every postings. Particularly you will not get any points:

1. If you post a response or a new submission without looking at earlier responses.
2. If you post a question with very little information.
3. If you post a question that is "VERY CLEARLY" due to skipping of reading material.

These DPPs in turn will help your grade. Note that 12 DPP is equivalent to one point for the final grade. If you get 120 DPPs, you will get the maximum point assignment (10 points) that one can get for participating in discussion forum. You may be able to earn extra credit of up to 6 points if you exceed 120 DPPs. The extra credit points will be determined by the relative participation performance of students in the discussion forum.
NOTE: I will like to discourage scenarios where most of the responses /comments are being posted by one or two students. If I see this happening, I will take necessary action.

**FAQ Related to Discussion Board**

Q1. What if I do not understand a topic?

Please post your question for clarification in the discussion board.

Q2. What if the instruction for completing the assignment is not clear?

Visit the Discussion forum and view the documents under the "Problems and Concerns" forum. If you do not find your answer there, please post your question in the discussion board under the forum "Problems and Concerns". Your instructor/TA will do his best to respond to your posted questions as quickly as possible. You will see the response posted in the discussion board.

Q3. What if the instruction for completing the assignment is clear, but I have no idea how to proceed?

In this course it is OK to discuss assignments including different approaches with other students in the discussion forum. However, it should be kept in mind that students should not post the code listing for assignments in the discussion forum. All students are expected to complete the assignment on their own.

Q4. If I am not able to finish my assignment because I am getting some error, which I do not understand, can I get some help from the TA or instructor?

No. You will not be provided any help by the instructor or TA for completing your assignment. However, you can post your question in the discussion board. After the due date, if you are still not able to resolve your problem, instructor or TA will help you in identifying the problem.

Q5. What is the role of instructor or TA in the discussion board? Is instructor going to post responses on student questions?

The main objective of the discussion board is to provide a platform where students can collaborate. The role of the Instructor is to monitor and assign points for participation in the discussion board. Instructor or TA will not respond to questions related to assignment completions. Questions specifically addressed to the instructor for clarification or concern will be responded by the instructor.

Q6. What kind of material I can post in the discussion board?

- Any problem and concern with the course that requires instructor attention.
- Post questions related to course that you would like to discuss with other students.
- Post responses to other student's questions

Q7. What is the best way of increasing my DPP (Discussion Participation Points)?

Read the questions posted by other students and try to respond to as many questions as you can. Though the correctness of your response is not required, however, please refrain from responding just for the sake of increasing your DPP. Also, if you have doubts about a particular topic, post your questions for clarifications in the discussion board.

**Schedule**

**Course Policies**

ple.odu.edu/courses/201920/cs724/print/0551eaa52ea6b33e363d04500f3859b07815dddb45a053b07867586ac23c22e20612342f?ref=%2Fcourses%…
Online Classroom Conduct (Netiquette)
Students are expected to follow good Netiquette rules. Netiquette is the accepted behavior for online participation. The following is a list of general guidelines for this course:

- Check your grammar and spelling
- Keep your comments focused on the topic
- Strive to write succinctly and clearly
- Share your knowledge and include supportive evidence for your comments
- Do not use all capital letters as that is viewed as shouting
- Avoid flaming—disrespectful language is unacceptable

Select the link to find more information on Netiquette.

Attendance
Since this is an on-line course, there is no mandatory attendance policy. However, students are expected to actively participate in the discussions, homework submissions, and journal writing. Each of these components is graded and counted towards the final grade.

Late submissions
If a situation has occurred that requires your time and attention, and will prevent submitting your work on time, please notify your instructor 24 hours before the scheduled due date. Otherwise, no late submissions will be allowed.

Course Disclaimer
Every attempt is made to provide a syllabus that is complete and that provides an accurate overview of the course. However, circumstances and events may make it necessary for the instructor to modify the syllabus during the semester. This may depend, in part, on the progress, needs, and experiences of the students.

University Policies

Honor Pledge
"I pledge to support the honor system of Old Dominion University. I will refrain from any form of academic dishonesty or deception, such as cheating or plagiarism. I am aware that as a member of the academic community, it is my responsibility to turn in all suspected violators of the honor system. I will report to Honor Council hearings if summoned."

By attending Old Dominion University you have accepted the responsibility to abide by this code. This is an institutional policy approved by the Board of Visitors. For more information please visit the Honor Council.

Special Needs
Old Dominion University is committed to achieving equal educational opportunity and full participation for persons with disabilities. It is the university's policy that no qualified person be excluded from participation in any university program or activity, be denied the benefits of any university program or activity, or otherwise be subjected to discrimination with regard to any university program or activity. This policy derives from the university's commitment to Non-discrimination for all persons in employment, access to facilities, student programs, activities and services. For additional information visit the Office of Educational Accessibility online or at 1525 Webb Center.

Disability Services.
In compliance with PL94-142 and more recent federal legislation affirming the rights of disabled individuals, provisions will be made for students with special needs on an individual basis. The student must have been identified as special needs by the university and an appropriate letter must be provided to the course instructor. Provision will be made
based upon written guidelines from the University's Office of Educational Accessibility. All students are expected to fulfill all course requirements.

University Email Policy

The Old Dominion University e-mail system is the official electronic mail system for distributing course-related Communications, policies, Announcements and other information. In addition, the University e-mail user ID and password are necessary for authentication and access to numerous electronic resources (online courses, faculty Web pages, etc.)

Withdrawal

A syllabus constitutes an agreement between the student and the course instructor about course requirements. Participation in this course indicates your acceptance of its teaching focus, requirements, and policies. Please review the syllabus and the course requirements as soon as possible. If you believe that the nature of this course does not meet your interests, needs or expectations, if you are not prepared for the amount of work involved - or if you anticipate that the class meetings, assignment deadlines or abiding by the course policies will constitute an unacceptable hardship for you - you should drop the class by the drop/add deadline, which is located in the ODU Schedule of Classes. For more information, please visit the Office of the Registrar.