Course Syllabus Part II
DSC 650 - Big Data
3 Credit Hours

Course Resources

Course Text(s)

Big Data: Principles and best practices of scalable realtime data systems (1st Edition).
Nathan Marz & James Warren
Manning Publications
ISBN-13: 978-1617290343
ISBN-10: 1617290343

Spark: The Definitive Guide: Big Data Processing Made Simple (1st Edition).
Bill Chambers & Matei Zaharia
O'Reilly Media
ISBN-13: 978-1491912218
ISBN-10: 1491912219

Required Resources

In this course, you will need to be able to:

- Access the Internet.
- Access Cyberactive.
- Access to Github
- Collaborate Online via Video and Voice.
- Collaborate while writing a single document.
- Submit a Word Document.
- Access to GitHub account.
- Python programming environment using PyCharm, Anaconda, and Jupyter Notebook.

Course Schedule

Week	Topic	Assigned Reading
1	Introduction to Big Data	*Big Data: Principles and best practices of scalable realtime data systems* - Chapters 1; *Spark: The Definitive Guide* - Chapters 1 and 2
2	Data Models	*Big Data: Principles and best practices of scalable realtime data systems* - Chapter 2; *Spark: The Definitive Guide* - Chapters 4 and 5
3	Data Storage	*Big Data: Principles and best practices of scalable realtime data systems* - Chapter 4; *Spark: The Definitive Guide* - Chapters 6 and 9

4	Batch Processing Fundamentals	*Big Data: Principles and best practices of scalable realtime data systems* - Chapter 6; *Spark: The Definitive Guide* - Chapters 7 and 8
5	Batch Processing Pipeline	*Big Data: Principles and best practices of scalable realtime data systems* - Chapter 8; *Spark: The Definitive Guide* - Chapters 10, 11, and 12
6	Realtime Views	*Big Data: Principles and best practices of scalable realtime data systems* - Chapters 12 and 13
7	Stream Processing	*Big Data: Principles and best practices of scalable realtime data systems* - Chapters 14 and 16; *Spark: The Definitive Guide* - Chapter 20
8	Structured Streaming	*Spark: The Definitive Guide* - Chapters 21 and 22
9	Analytics and Machine Learning	*Spark: The Definitive Guide* - Chapters 24 and 25
10	Classification and Regression	*Spark: The Definitive Guide* - Chapters 26 and 27
11	Recommendations and Graph Analytics	*Spark: The Definitive Guide* - Chapters 28 and 30
12	End-to-End Big Data Architecture	*Big Data: Principles and best practices of scalable realtime data systems* - Chapter 18; *Spark: The Definitive Guide* - Chapters 15, 16 and 17

Course Activities

In this section of the syllabus, I will describe what we will be doing in each of the activities for each week. Specifically, I will be describing your deliverables – those items you need to submit at or before the deadline. You can find more detail on grading criteria for each category by viewing its detailed rubric.

Written Assignments

In some weeks, you will be assigned a written assignment aligning to the weekly reading and topics, which is due to the discussion board. This post must be 1000 words minimum and contain at least two credible sources. It should be written with an introduction, body and conclusion and cited references.

Exercises

Every other week, you will be assigned an exercise or series of exercises based on the weekly topic to complete and submit to the assignment link. These are not group assignments to complete and should be done on your own. However, if you have questions about a specific exercise, you are encouraged to use the discussion board to discuss with your classmates, without completing the assignment together.

Discussion

Each week, you will be making 2 discussion posts in the specified forums. These two posts can either be responses to a fellow classmate or they can be something you found interesting in the reading/homework or something you did not understand or agree with.

Each post must be a minimum of 250 words and contain at least one credible source. These responses should be "substantive" which means more than, "Neat!" or "Good job!" They should also not contain jargon or be a post that boils down to you reposting the same thing you're commenting on in a different way.

Grade and Point Breakdown

Component:	Percentage	Point Value	Number of Times	Total
Discussion	33.3%	50 Points	2 Times per Week for 12 Weeks	600
Written Assignments	33.3%	100 Points	6	600
Exercises	33.3%	100 Points	6	600
			Total Points	1800

Late Work

Late work is not accepted unless arrangements are made with the instructor for very special, unavoidable circumstances. If you do not alert the professor before or shortly after something that will make you late, the chances of special arrangements are much lower. If in doubt, please email as soon as possible.

Participation

Students are expected to login often and contribute to the class on a regular basis, including posting to the discussion board, submitting assignments, and participating in group activities as required. If you have specific participation requirements related to your educational funding or student status, you are expected to monitor your own participation to ensure you are in compliance with those requirements.

Expectations for Students

- Students should expect to spend approximately 10-15 hours per week to complete the activities and assignments in this course.
- Students will log in as often as needed to complete their assignments and progress through the course.
- Students will treat their classmates and the instructor with respect and courtesy.

- Students are responsible for keeping current with the reading assignments and coming to class prepared to discuss the work assigned.
- Students are responsible for knowing what assignments are due and when.
- Students will submit only their own work and will not commit plagiarism or other acts of academic dishonesty.
- Students will contact the instructor as soon as personal problems arise that may affect the student's ability to complete assignments on time.

Expectations for Faculty

- The instructor will treat all students with respect and courtesy.
- The instructor will make grading criteria clear and follow the criteria scrupulously in evaluating student work.
- The instructor will provide feedback about student work within 6 days of due dates (or 24 hours prior to the next due date)—feedback that helps the student learn and improve.
- The instructor will respond to all student messages within 48 hours.