Course Syllabus Part II

DSC 650: Big Data

3 Credit Hours

Course Resources

Course Text(s):

1. **Big Data: Principles and best practices of scalable realtime data systems** by Nathan Marz and James Warren. Manning Publications; 1st edition. ISBN-13: 978-1617290343
2. **Hadoop: The Definitive Guide** by Tom White. O'Reilly Media; 4th edition. ISBN-13: 978-1491901632
3. **Learning Spark: Lightning-Fast Data Analytics** by Jules S. Damji, Brooke Wenig, Tathagata Das, and Denny Lee. O'Reilly Media; 2nd edition. ISBN-13: 978-1492050049

Required Resources:

- Internet Access for research and accessing various big data platforms.
- Access to GitHub for code repository and project management.
- Ability to create and submit documents in Word, Excel, or PowerPoint formats.

Course Schedule

Week	Topic	Reading Assignment
1	Introduction to Big Data and Infrastructure	Big Data: Principles: Chapters 1-2 *White Paper (linked in the course)*
2	Hadoop Basics and MapReduce	Hadoop: The Definitive Guide: Chapters 5-6, Learning Spark: Chapters 3-4 *White Paper (linked in the course)*
3	Hive and Data Warehousing	Big Data: Principles: Chapters 4-5 *White Paper (linked in the course)*
4	Spark Fundamentals	Learning Spark: Chapters 6-7 *White Paper (linked in the course)*
5	Advanced Spark Programming	Hadoop: The Definitive Guide: Chapters 8-9 *White Paper (linked in the course)*
6	HBase and NoSQL Databases	Big Data: Principles: Chapter 7 *White Paper (linked in the course)*
7	Kafka and Real-time Data Streaming	Hadoop: The Definitive Guide: Chapter 12 *White Paper (linked in the course)*
8	Solr and Search Analytics	*White Paper (linked in the course)*
9	Nifi and Data Flow Management	*White Paper (linked in the course)*
10	Big Data Architectures	*White Paper (linked in the course)*
11-12	Big Data Project	

Effective Date of Syllabus: November 27, 2023

Course Activities

Discussion/Participation

Students are required to engage actively in online discussions. A minimum of 10 posts are expected weekly on the platform, fostering a vibrant learning community. The posts could be queries, insights, sharing relevant articles or resources, or responding to peers.

Exercises

Weekly exercises will focus on hands-on experience with various components of big data technologies. These exercises will involve working on real-world datasets and implementing solutions based on the concepts learned during the lectures.

Term Project

During the final two weeks of the course, students will have the opportunity to showcase the skills and knowledge they've acquired throughout the course. Students can select their preferred big data technologies to implement a real-world big data use case. The project aims to foster creativity and critical thinking, encouraging students to solve complex problems using the appropriate big data solutions.

Point Breakdown

Component	Percentage	Point Value Each Week	Number of Times	Total
Discussion/Participation	30%	60 Points	10 Posts Per Week	720
Exercises	40%	96 Points	Weekly	960
Term Project	30%	720 Points	Once	720
				Total Points 2400

Late Work

Late submissions are generally not accepted. In exceptional circumstances, reach out to the instructor beforehand to discuss potential accommodations.

Participation

Regular participation in the course is expected, including active involvement in discussions, timely assignment submissions, and engagement in group activities.

Expectations for Students

- Dedicate approximately 12-17 hours weekly for course-related activities.
- Maintain a respectful and courteous demeanor in all interactions.
- Stay abreast with the schedule and prepare accordingly for classes.

- Uphold academic integrity in all submissions.

Expectations for Faculty

- Facilitate a respectful and inclusive learning environment.
- Adhere to grading criteria and provide constructive feedback within 6 days of submission.
- Respond to student queries within 48 hours.
- Offer guidance and support for project development and execution.