Course Syllabus

Course Title: DS 610 Big Data Analytics
Course Prefix, Number and Section:
Professor:
Term:
Meeting Times:

Professor's Contact Information

Office Phone:
Department Phone:
Office Location:
Office Hours:
Email:

General Course Information

Course Objectives and Description

Big Data (Structured, semi-structured, & unstructured) refers to large datasets that are challenging to store, search, share, visualize, and analyze. Gathering and analyzing these large data sets are quickly becoming a key basis of competition. This course explores several key technologies used in acquiring, organizing, storing, and analyzing big data. Topics covered include Hadoop, unstructured data concepts (key-value), Map Reduce technology, related tools that provide SQL-like access to unstructured data: Pig and Hive, NoSQL storage solutions like HBase, Cassandra, and Oracle NoSQL and analytics for big data. A part of the course is devoted to public Cloud as a resource for big data analytics. The objective of the course is for students to gain the ability to employ the latest tools, technologies and techniques required to analyze, debug, iterate and optimize the analysis to infer actionable insights from Big Data.

Prerequisites

Curiosity and drive to learn

Learning Outcomes

At the completion of this course, students will be able to:

- Analyze Big Data with Apache Spark.
- Analyze Big Data with Apache Hadoop.
- Extract meaningful insights out of raw data using NoSQL databases.

Textbook and Readings

Spark: The Definitive Guide: Big Data Processing Made Simple by Bill Chambers, Matei Zaharia

ISBN-13: 978-1491912218

Course Components

Class Participation

Class participation will count for 20% of your grade. This includes weekly discussion posts and responses.

Assignments

Assignments will be worth 40% of your grade.

Midterm Exam

The midterm exam is worth 20% of your grade.

Final Exam

The final exam is worth 20% of your grade.

Assignments and Academic Calendar

[Topics, Homework Assignment Due Dates, Quiz Dates, Reading]

Week	Dates	In-Class or Online	Topic/Homework/Quiz	Due Date	Reading

1			Course Introduction & Python Recapitulation Week 1 Discussion Week 1 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 1 Lecture
2			Python with Spark Setup & Spark DataFrame Basics Week 2 Discussion Week 2 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 2 Lecture Read http://spark.apache.org/docs/latest/sql-programming-guide.html
3			Linear & Logistic Regression Week 3 Discussion Week 3 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 3 Lecture Read http://spark.apache.org/docs/latest/api/python/pyspark.sql.html
4			Decision Trees & Random Forests Week 4 Discussion Week 4 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 4 Lecture
5			Midterm Exam	Midterm Exam is due Sunday 11:59 pm	

6			Spark Streaming With Python Week 6 Discussion Week 6 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 6 Lecture
7			Data Ingestion & Extraction with Spark Week 7 Discussion Week 7 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 7 Lecture
8			HBase Essentials Week 8 Discussion Week 8 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 8 Lecture
9			Big Data Analytics with Hadoop I Week 9 Discussion Week 9 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 9 Lecture

10			Big Data Analytics with Hadoop II Week 10 Discussion Week 10 Assignment	Discussion is due Wednesday 11:59 pm Comments due Sunday 11:59pm Assignment is due by Sunday 11:59pm	Week 10 Lecture
11			Final Exam	Final Exam is due Sunday 11:59 pm	

Grading Criteria and Weighting

Assignments and Grade Breakdown

Categories	**Weights**
Participation & Quizzes	20%
Assignments	40%
Midterm Exam	20%
Final Exam	20%
TOTAL	100%

Grading Scale

Grade	**Performance**	**Numeric Grade**	**Grade Point**
A	Outstanding	94 to 100	4.0
A-	Excellent	90 to 93	3.7
B+	Very Good	87 to 89	3.3
B	Good	83 to 86	3.0
B-	Above Average	80 to 82	2.7
C+	Average	77 to 79	2.3
C	Satisfactory	70 to 76	2.0
F	Failure of Course	59 or below	0.0

Course Policies

Class Attendance- Absences and Lateness

According to Saint Peter's policy, a student is allowed four absences in a term for a course that meets two times a week. If you exceed this number, you may receive an FA for failure due to excessive absences. Arriving on time is expected. It is not appropriate to leave the class during the period except for emergencies.

Class Technology

Technology used in this course and in the classroom includes email, Blackboard, PowerPoint, and programs such as Word.

Cell Phones, PDAs, Other Devices (Classroom Etiquette)

Cell phones, personal digital assistants, and other devices such as laptops should be turned off in the classroom.

Online Etiquette and Anti-Harassment Policy

The use of University sponsored emails, portals as well as information and material accessed on the Saint Peters network should be in keeping with University values and Student Code of Conduct. University guidelines for responsible use of technology can be found in the Student Handbook.

Academic Integrity

Violations of professional ethics will not be tolerated. This includes plagiarism, cheating, or false attendance. If a student is suspected of such behavior, standard University policy will be used to deal with the specific situation (see the Student Code of Conduct). It is considered plagiarism to use online sources, texts, other students' work, etc. as your own work, so you MUST quote and cite all your references and put phrases in your own words to avoid unintended plagiarism of Internet and other sources.

Accommodative Services

Please let me know at the beginning of the semester about any learning accommodations you need, and make sure you have the appropriate paperwork from the Academic Dean's Office. Saint Peter's Faculty is responsible for providing access to education which is free from discrimination. Students apply for academic accommodations by submitting the appropriate forms to the Center for Academic Success and Engagement. Academic accommodations are approved based on a student's individualized needs. For more information please visit the Accommodations and Services webpage.

Student Support Services

If you need any additional help, please contact TRiO Student Support Services.

Title IX Compliance

In the event that you choose to write or speak about having survived sexualized violence, including rape, sexual assault, dating violence, domestic violence, or stalking, Saint Peter's University Policies require that, as your instructor, I share this information with the Title IX

Coordinator, Elena Serra. Elena or a trained member of her team will contact you to let you know about support services at Saint Peter's as well as options for holding accountable the person who harmed you. Whereas you are not required to speak with them, they will share resources with you.

Instructional Continuity Plan

For classes that normally entail meeting face-to-face, the continuity plan will be in place by including some synchronous course activities if access to campus is restricted for more than one week. Also, the use of Blackboard is mandatory and the only LMS that should be used for all SPU classes is SPU's Blackboard.

Evaluations