CIIC 5140 - Big Data Analytics

1. General Information:

Alpha-numeric codification: CIIC 5140
Course Title: Big Data Analytics
Number of credits: 3
Contact Period: 3 hours per week

2. Course Description:

English: Description of the principles of big data systems and analysis techniques for the design of cloud computing processes. Discussion of the implementation of parallel algorithms to process data on cloud-resident storage and memory-based file systems.

Spanish: Descripción de los los principios de sistemas datos a gran escala (*big data*) y técnicas de análisis para el diseño de procesos de computación en la nube. Discusión de la implantación de algoritmos paralelos para procesar datos en sistema de almacenaje en la nube, y en archivos de memoria principal.

3. Pre/Co-requisites and other requirements:

Pre-requisites: (CIIC 4060 or ICOM 5016) and ININ 4010 or authorization of department chair.

4. Course Objectives:

The fundamental techniques to model and solve big data problems will be explained, using software tools and algorithms that can be scaled to large datasets and operated on loosely coupled cluster/cloud platforms.

5. Instructional Strategies:

☒conference ☐discussion ☒computation ☐laboratory

☐seminar with formal presentation ☐seminar without formal presentation ☐workshop

☐art workshop ☐practice ☐trip ☐thesis ☐special problems ☐tutoring

☐research ☐other, please specify:

6. Minimum or Required Resources Available:

Students will use the Departmental computer laboratories to complete course projects.

7. Course time frame and thematic outline

Outline	Contact Hours
Introduction	1
Map Reduce	3
Google and Hadoop File Systems	3
Reliable cluster management	3
Hive Data Warehouse	4
Map Reduce Query Processing	6
Memory Resilient Data Sets	3
Spark System	4
Association Rules, Classification, Decision Trees, K-means clustering	9
Graph Analytics	6
Stream Processing	3
Partial Exams	3
Total hours: (equivalent to contact period)	45

8. Grading System

☒Quantifiable (letters) ☐ Not Quantifiable

9. Evaluation Strategies

	Quantity	Percent
☒ **Exams**	2	25%
☒ **Final Exam**	1	20%
☒ **Short Quizzes**	5-10	15%
☐**Oral Reports**		
☐ **Monographies**		
☐ **Portfolio**		

☒ **Projects**	1	40%
☐**Journals**		
☐**Other, specify:**		
TOTAL:		**100%**

10. Bibliography:

1. Tom White, *Hadoop: The Definitive Guide,* 4th Edition, O'Reilly Media, 2015
2. Sandy Ryza, Uri Laserson, Sean Owen, and Josh Willis, *Advanced Analytics with Spark: Patterns for Learning from Data at Scale,* O'Reilly Media, 2015
3. Cathy O'Neil, and Rachel Schutt, *Doing Data Science: Straight Talk from the Frontline,* O'Reilly Media, 2013
4. A collection of papers from recent Big Data Conferences, as determined by the instructor.

11. Course Outcomes

After successfully completing the course, the student will be able to:	Program Student Outcomes Impacted
1. distinguish the criteria to classify a problem as a big data problem based on data volume, speed of generation, data diversity, data veracity, and data variability.	7
2. apply data processing techniques to handle big data sets, including: map-reduce, direct acyclic graph (DAG) processing, resilient distributed datasets (RDD), data streaming, and graph processing.	1
3. apply parallel data management and data warehousing techniques to solve big data problems.	2
4. design, analyze, and implement machine learning techniques atop big data systems.	2
5. build big data applications using contemporary big data programming frameworks.	6

12. According to Law 51

Students will identify themselves with the Institution and the instructor of the course for purposes of assessment (exams) accommodations. For more information please call the Student with Disabilities Office which is part of the Dean of Students office (Office #4) at (787)265-3862 or (787)832-4040 extensions 3250 or 3258.

13. Academic Integrity

-The University of Puerto Rico promotes the highest standards of academic and scientific integrity. Article 6.2 of the UPR Students General Bylaws (Board of Trustees Certification 13, 2009-2010) states that academic dishonesty includes, but is not limited to: fraudulent actions; obtaining grades or academic degrees by false or fraudulent simulations; copying the whole or part of the academic work of another person; plagiarizing totally or partially the work of another person; copying all or part of another person answers to the questions of an oral or written exam by taking or getting someone else to take the exam on his/her behalf; as well as enabling and facilitating another person to perform the aforementioned behavior. Any of these behaviors will be subject to disciplinary action in accordance with the disciplinary procedure laid down in the UPR Students General Bylaws.—