ACADEMIC PROGRAM

DATA PROCESSING

B.F.A. IN **COMPUTER SCIENCE**

MODALITY: ON CAMPUS

ACADEMIC YEAR: 2023-2024

Name of the course:	Data Processing	
Degree :	Computer Science	
Location:	Centro Universitario de Tecnología y Arte Digital	
Area:	Data Engineering	
Year:	3º	
Teaching period:	2	
Type:	OBM	
ECTS credits:	6	
Teaching modality:	On campus	
Language:	English	
Lecturer / Email	-	
Web page:	http://www.u-tad.com/	

SUBJECT DESCRIPTION

Area description

The contents of the subject allow students to understand the flow of searching, ingesting, storing, processing and analyzing data information and brings students closer to the techniques and technologies necessary for managing large amounts of data.

Subject description

The objective of this course is to enter the world of Big Data in a distributed environment and in real time.

Learning distributed data processing is essential today, in this course we will use Spark. Spark is the technology that is revolutionizing the world of analytics and big data. Spark is an open source data processing engine built around speed, ease of use, and analytics.

We will start the course with an Introduction to Scala with the goal of using Spark using SparkShell. We will cover basic Spark concepts such as distributed tasks, RDD, and the Master/Slave architecture and then explore Spark SQL modules/extensions, Spark Streaming, and Machine Learning with MLib in Spark.

COMPETENCIES AND LEARNING OUTCOMES

Competencies

BASIC AND GENERAL SKILLS

CG1 - Ability to understand, schedule and solve problems trough software development

CG3 - Knowledge of the scientific fundamentals applicable to the resolution of computer problems

CG4 - Ability to simplify and optimize computer systems by understanding their complexity

CG9 - Ability to learn, modify and develop new software solutions

CG10 -Use of creative techniques to carry out computer projects

CG11 - Ability to search, analyze and manage information for insights capture

BC1: Students should demonstrate knowledge in an area of study that builds upon the foundation of general secondary education and goest beyond at a level that, while supported by advanced textbooks, also encompasses certain aspects derived from the cutting edge of their field of study.

BC2: Students should be able to apply their knowledge to their work or vocation in a professional manner, and they should possess the competencies typically demonstrated through the development and defence of arguments as well as problem-solving within their field of study.

BC3: Students must possess the ability to gather and interpret relevant data (usually within their field of study) in order to make judgments that involve reflection on socially, scientifically, or ethically significant issues.

BC4: Students should be capable of conveying information, ideas, problems, and solutions to both specialized and non-specialized audiences.

BC5: Students should have developed the learning skills necessary to pursue further studies with a high degree of autonomy.

TRANSVERSAL SKILLS

CT1 - Knowledge of the definition, scope and implementation of the fundamentals of project management methodologies for technology projects

CT2 - Knowledge of the main sectorial players and the life cycle of a digital content development and commercialization project

CT4 -Ability to update the knowledge acquired in the management of digital tools and technologies according to the current state of affairs of the sector and the technological solution

CT5 -Development of the basic skills for digital entrepreneurship.

SPECIFIC SKILLS

CE3 - Knowledge of relational algebra and querying in procedural languages for the design of standardized database schemas based on entity-relationship models

CE10 - Ability to work with a release manager and generate application documentation automatically.

Learning outcomes

Upon completion of the degree, the graduate will be able to:

- To know and develop storage procedures and data management in distributed environments.

- To know and apply supervised, unsupervised and semisupervised learning techniques.

- To know and apply Deep Learning techniques

- To be able to retrieve information using web scraping or standard APIs

- To know and understand Natural Language Processing techniques

- To be able to analyze social networks contents.

- To understand the nature and representation of digital images.

- To know the applications of neural networks to the analysis and generation of sound, static images and video.

- To develop software solutions for computer vision.

- To develop a fully-fledged data project applying iterative methodology from design to delivery.

CONTENTS

Data distribution and fragmentation

Distributed concurrency

Check and reliability protocols

Data Replication management

Distributed data management architectures

SUBJECT SYLLABUS

Topic 1.- Introduction to Spark and Scala

Introduction and installation of Spark. Introduction to data structures in Scala: lists, dictionaries and data frames. Structure manipulation methods: understood list, anonymous/lambda functions and via map/reduce/filter, apply and fold.

Topic 2.- Basic Spark

Spark basics. Spark Core, distributed tasks, programming and basic I/O and RDD (Resilient Distributed Datasets) functionalities.

Topic 3.- Spark Cluster

Master/Slave architecture - slaves/workers in the case of Spark. The controller and executors run their individual processes and users can run them on the same Spark cluster or on separate machines.

Topic 4.- Spark SQL

Spark SQL is a Spark module for structured data processing. It provides a programming abstraction called DataFrames and can also act as a distributed SQL query engine.

Topic 5.- Spark Streaming

Spark Streaming is an extension to the core Spark API that enables scalable, high-performance, fault-tolerant stream processing of real-time data streams. Spark Streaming provides a high-level abstraction called discretized stream or DStream, which represents a continuous stream of data.

Topic 6.- Machine Learning with MLib in Spark

MLlib is Spark's machine learning (ML) library. Its goal is to make practical machine learning scalable and easy. At a high level, it provides tools such as: ML Algorithms: Common learning algorithms such as classification, regression, clustering, and collaborative filtering.

Topic 7.- Graphs with GraphX in Spark

GraphX is a distributed graph processing library that is part of the Apache Spark ecosystem. This library provides a set of specific abstractions and operations for working with large-scale graph data in an efficient and distributed manner.

TRAINING ACTIVITIES AND TEACHING METHODOLOGIES

TRAINING ACTIVITIES

LEARNING ACTIVITIES	Total hours	Hours of presence
Theoretical / Expository classes	29,38	29,38
Practical classes	23,25	23,25
Tutorials	4,00	2,00
Independent study and autonomous work of the student	50,00	0,00
Elaboration of work (group or individual)	31,88	0,00
Evaluation Activities	5,25	5,25
Project Follow-Up	6,25	6,25
TOTAL	150	66,13

Teaching methodologies

Expository method or master lesson

Case learning

Learning based on problem solving

Project based learning

Cooperative or collaborative learning

inquiry learning

Flipped classroom methodology

Gamification

Just in time Teaching (JITT) or classroom on time

Expository method or master lesson

Case method

Learning based on problem solving

Project based learning

Cooperative or collaborative learning

inquiry learning

Flipped classroom methodology

Gamification

TEMPORAL DEVELOPMENT

DIDACTIC UNITS / TOPICS TIME PERIOD

Introduction to the subject, virtual machine installation Week 1

Introduction to scala (structures and functions, lambda and map/reduce/filter) Week 2

Spark Basic Weeks 3 to 5

Spark Cluster Weeks 6 to 7

Spark SQL Weeks 8 to 9

Partial Weeks 10

Spark Streaming Weeks 11 to 12

Machine Learning with MLib in Spark Weeks 13 to 14

Spark GraphX Weeks 15

EVALUATION SYSTEM

ASSESSMENT SYSTEM	MINIMUM SCORE RESPECT TO THE FINAL ASSESSMENT (%)	MAXIMUM SCORE RESPECT TO THE FINAL ASSESSMENT (%)

Assessment of participation in class, exercises or projects of the course	10	30
Assessment of assignments, projects, reports, memos	40	80
Objective test	10	60

GRADING CRITERIA

ASSESSMENT SYSTEM	ORDINARY EVALUATION	EXTRAORDINARY EVALUATION
Assessment of participation in class, exercises or projects of the course	10	10
Assessment of assignments, projects, reports, memos	40	40
Objective test	50	50

General comments on the evaluations/assessments

The final grade for the subject in the ordinary session will be calculated with the grades of the three activities described in the previous table, applying the corresponding weights to them. Therefore, the rule to apply to calculate the final grade will be:

Ordinary_Final_Grade = 10%*AE1 + 40%*AE2 + 50%*AE3

- AE1: The evaluation of participation and class work: will be carried out based on attendance, class work and the delivery of the activities and exercises proposed during classes. This aspect will represent 10% of the final grade for the subject in the ordinary call.

When should I make the extraordinary call and its conditions:

5. Students with less than 4 in the ordinary final exam but with the mandatory practical activities (those in group AE2) presented and approved must appear exclusively for the extraordinary call exam and the rest of the course grades will be kept. Your average grade for the extraordinary call will therefore be:

Final grade for extraordinary call = 10% attendance and class work

 + 40% of practice(s) and partial + 50% Extraordinary exam

• If the student has the subject failed in ordinary due to pending and/or failed mandatory practices but his exam grade is greater than or equal to 4, said exam is saved, but he would have to present the pending practice(s) or suspended in extraordinary call.

Final grade for extraordinary call = 10% attendance and class work

+ 40% of practice(s) conv. extraordinary + 50% Ordinary Exam

6. The rest of the students who have failed the subject must take an extraordinary final exam and deliver in an extraordinary manner the practical(s) that the teacher has defined and requested for completion during Q2, their final grade being:

Final grade for extraordinary call = 10% attendance and class work

+ 30% practice(s) conv. extraordinary + 60% Extraordinary Exam Note

- AE2: Throughout the course, mandatory practice(s) will be proposed that must be delivered on the indicated date through the virtual platform. Additionally, a partial exam of the subject taught so far will also be carried out. This activity (mandatory and partial practice(s)) will be evaluated through the virtual platform itself and will account for 40% of the final grade for the subject in the ordinary call. Works out of form and date will not be admitted without justified cause and all of them must be submitted to be approved in the ordinary call. If any of these mandatory evaluable practices are delivered after the deadline, there will be a 30% penalty on the grade. In order to pass the subject in the ordinary session, it will be a requirement to have delivered the mandatory practice(s) that have been requested. If the student has not completed the proposed midterm, they will have a 0 in said grade.

- AE3: At the end of the semester a final exam will be taken, which will have a total weight on the grade of the ordinary call of 50%. To pass the subject in the ordinary session, it will be necessary for the student to have at least a 4 (out of 10) in said exam.

- In summary, to pass the subject in the ordinary call it is necessary that the final exam be at least 4.0 (out of 10), that the evaluable practical(s) requested during the course be delivered (those belonging to activity AE2 of the 40%), that is, these practices are mandatory and no average will be taken if any of them is not presented and the final average grade (including AE1, AE2 and AE3) is at least 5.0 (out of 10).

If any of these requirements are not met, the subject will be automatically considered failed regardless of the rest of the grades.

- If the student does not pass the ordinary call, he or she may take the extraordinary call. Namely:

• IF the student has failed the subject in the ordinary session because he did not reach 4 in the final exam, but has all his practices delivered and approved, then he will have to take the extraordinary exam exclusively and the rest of his grades from the exam will be kept. course. Her extraordinary exam maintains the weight of 50% of the grade and must have a minimum grade of 4 (out of 10).

• If the student has the subject failed in ordinary due to pending and/or failed practices but his exam grade is greater than or equal to 4, said exam will be saved, but he will have to present the pending or failed practices in an extraordinary session.

• Otherwise, (failed exam and failed mandatory practices) you must:

• Take the extraordinary final exam that will represent 60% of the grade in said call and in which the subject required of the student will be all the content of the subject seen in class.

• In addition, you must submit the practice(s) failed or not submitted and that are presented at the beginning of the second semester and whose weight on the final grade will be 30%.

• The remaining 10% will be the grade for attendance and class work obtained during the course.

The final average grade must be at least 5.0 (out of 10) and the final exam grade must be at least 4 (out of 10).

- No grades of any kind will be kept between different academic years, nor between different calls.

Ordinary Evaluation Details

How do I approve the ordinary call:

2. To pass in the ordinary call, the student must have a grade equal to or greater than 5.00 in the final average of all grades (including the practice grade(s) and partial grade, the final exam grade and the attendance and class work).

3. In order for this average to be taken, the student must have at least a 4.00 in the Final Exam and have delivered the mandatory practice(s) requested during the course.

4. In order for the obligatory evaluable practical activities to be graded, it will be necessary to have completed the deliveries on the date established by the teacher, that is, work out of form and date will not be admitted without justified cause. Unjustified delays in deliveries will have a 30% grade penalty.

Extraordinary call details

When should I make the extraordinary call and its conditions:

5. Students with less than 4 in the ordinary final exam but with the mandatory practical activities (those in group AE2) presented and approved must appear exclusively for the extraordinary call exam and the rest of the course grades will be kept. Your average grade for the extraordinary call will therefore be:

Final grade for extraordinary call = 10% attendance and class work

  + 40% of practice(s) and partial + 50% Extraordinary exam

• If the student has the subject failed in ordinary due to pending and/or failed mandatory practices but his exam grade is greater than or equal to 4, said exam is saved, but he would have to present the pending practice(s) or suspended in extraordinary call.

Final grade for extraordinary call = 10% attendance and class work

+ 40% of practice(s) conv. extraordinary + 50% Ordinary Exam

6. The rest of the students who have failed the subject must take an extraordinary final exam and deliver in an extraordinary manner the practical(s) that the teacher has defined and requested for completion during Q2, their final grade being:

Final grade for extraordinary call = 10% attendance and class work

+ 30% practice(s) conv. extraordinary + 60% Extraordinary Exam Note

LIST OF REFERENCES (BOOKS, PUBLICATIONS, WEBSITES):

Basic Bibliography:

Holden Karau, Learning Spark: Lightning-Fast Big Data, 2015

Recommended Bibliography:

AI and Big Data online book: https://iaarbook.github.io/

REQUIRED MATERIALS, SOFTWARE AND TOOLS

Type of classroom
Theory classroom

Board and projection system

Materials:
Personal Computer

Software:
Virtual Box

Ubuntu 22.04