



COURSE SYLLABUS

I. Course Information :

Course Code	
Course Name	Big Data
Department	Data Science
Major	Data Science
Academic Year	2022-2023
Semester Duration	Spring
Number of Credits	
Number of Hours	50
Instructor Name	Dr. Malak Khreis
Email	
Phone number	
Office Hours	

II. Course Description

Big Data can take up terabytes and petabytes of storage space in diverse formats including text, video, sound, images, and more. The course gives an overview of the Big Data phenomenon, focusing then on extracting value from the Big Data using predictive analytics techniques.

III. Course Objectives

This course aims to provide students with/allows students to:

- Describe the Big Data landscape including examples of real world big data problems and approaches.
- Identify the high level components in the data science lifecycle and associated dataflow.
- Explain the V's of BigData and why each impacts, monitoring, storage, analysis and reporting, including their impact in the presence of multiple V's.
- Identify bigdata problems and be able to recast problems as data science questions.
- Summarize the features and significance of the HDFS file system and the MapReduce programming model and how they relate to working with BigData.
- Understanding the architecture of Hadoop including Sqoop, Kafka and Hbase.

In order to achieve the objectives of the course, the following requirements must be met by the student:



- Scientific activities
- Academic regulations
- Academic work and perseverance
- 70% attendance and participation

IV. Course Outline and Calender

Module	Key Concepts	Assignments
Introduction to Big Data	• Introduction to Big Data • Concepts of Big Data • Where does big data come from? • Characteristics of Big data: Velocity, Volume, Variety • Facts about Big Data • Getting Value out of Big data • The Process of data Analysis • Big Data Ecosystem Overview	
Cloud computing, and big data interaction	• Introduction: Know what economic advantages are driving the adoption of Cloud Computing • Virtualization, cloud computing, and big data interact to create exciting business opportunities • Principal architectural components of a cloud and their organization • Services and orchestration on cloud: IaaS, PaaS, and SaaS • Compare IaaS, PaaS, and SaaS • Services that major Cloud companies provide and how they provide them	



Apache Hadoop	• HDFS ○ HDFS Architecture ○ NameNode ○ SecondaryNameNode ○ DataNode ○ Configuring Hadoop Environment on Ubuntu ○ Different modes of installation ○ Pseudo Distributed Mode ○ Standalone mode ○ Fully Distributed Mode ○ HDFS Vs Regular File System	Preparing Big Data Environment
MapReduce	• MapReduce ○ Introduction to MapReduce ○ MapReduce Engine ○ JobTracker ○ TaskTracker ○ MapReduce ProgrammingModel ○ Mapper Class	Hadoop MapReduce

P.S.: The course outline above may be modified during the semester.

V. Teaching approach & methods

- Lectures: power point
- Cloudera Platform for application

VI. Assessment Methods

The success of the student in this course is based on:

- 1- His/her active participation in class.
- 2- His/her scientific contribution through research done outside the classroom.
- 3- His/her reading notes, research reports and practical applications as requested by course instructor.

The assessment of the student is carried out according to the following grid:

Assessment Type	Final Grade Weighting (%)	Dates/Deadlines
Participation (projects, research, oral presentation)	30%	



Homework (Hadoop)	10%	
Final exam	60%	
Total	100%	

VII. Specific Course Requirements

1- Regulations concerning deadlines for written and applied work submissions

The teacher has the right to:

- Refuse to receive written applied work outside the deadlines.
- Give the student additional time after omitting up to 25% of the mark awarded to the concerned activity.

2- Academic measures in case of plagiarism

The student shall be liable to the following penalties in case of any attempt of illicit scientific citation:

- Cancellation of the student's contribution, activity or scientific article and reduction of the final mark.
- If the same offense is reported in other courses, cancellation of the final exam and failure in the relevant course(s).

VIII. References

--The Big Data Textbook, Teaching large-scale databases in universities, by Ghislain Fourny, (2016-2023).

-- An Introduction to Big Data Concepts and Terminology, By Justin Ellingwood, 2016.

-- An Introduction to Big Data Concepts, By: Rancher Admin, 2019.

--Big Data: Principles and best practices of scalable realtime data systems First Edition, by Nathan Marz , James Warren.

--<https://hive.apache.org/>

--<https://cwiki.apache.org/confluence/display/Hive/Presentations>

--<https://developer.yahoo.com/blogs/hadoop/comparing-pig-latin-sql-constructing-data-processing-pipelines-444.html>

--<http://www.qubole.com/blog/big-data/hive-best-practices/>

--https://issues.apache.org/jira/secure/attachment/12411185/hive_benchmark_2009-06-18.pdf