



COLLEGE OF INFORMATION TECHNOLOGY
DEPARTMENT OF INFORMATION TECHNOLOGY
COURSE SYLLABUS/ SPECIFICATION

Course Code & Title: ITCS 520 – Big Data Analytics

Weight: (3 - 0 - 3)

Prerequisite: ITCS 511

NQF Level Allocated: 9

NQF Notional Hours / Credits: 120 notional hours/ 12 NQF credit

Description: This course covers foundational techniques and tools required for data science and big data analytics. The course focuses on concepts, principles, and techniques applicable to any technology environment and industry with emphasis on systems and algorithms for large-scale advanced data analysis. Topics covered include concepts and algorithms for building big data systems, data analytics lifecycle, basic and advanced analytics methods, and emerging big data technology and tools.

Objective:

1. To explain advanced big data analytics concepts, techniques and methods.
2. To teach students how to apply big data analytics lifecycle and use advanced tools to address real business problems.
3. To explain and demonstrate the effective use of emerging big data technologies such as Hadoop and MapReduce

Semester: First 2019-2020

Instructor (s): Dr. Karim Hadjar

Office Telephone:

Email (s): khajjar@ahlia.edu.bh

Intended Learning Outcomes (ILOs):

A. Knowledge and Understanding		NQF Descriptor/ Level
A1	Concepts and Theories: Demonstrate extensive knowledge and critical understanding of key concepts and theories of big data analytics.	Knowledge: theoretical understanding [Level 9]
A2	Contemporary Trends, Problems and Research: Demonstrate an informed and critical awareness of the current trends and advancements in big data analytics techniques, technologies and applications.	Knowledge: theoretical understanding [Level 9]
A3	Professional Responsibility: Demonstrate knowledge and understanding of techniques used to preserve appropriate level of privacy and security needed in big data systems.	Knowledge: theoretical understanding [Level 9]

B. Subject-specific Skills		NQF Descriptor/ Level
B1	Problem Solving: Develop solutions for business-related problems by applying the big data analytics lifecycle including discovery and formalization, data preparation, planning, implementation and results communication.	Skills: Communication, ICT and Numeracy Skills [Level 9] Knowledge: Practical Understanding [Level 9] Competence: Context [Level 8]
B2	Modeling and Design: Develop appropriate models for big data systems and visualization of analytics results to convey clear insights.	Skills: Communication, ICT and Numeracy Skills [Level 9]

		Knowledge: Practical Understanding [Level 9] Competence: Context [Level 8]
B3	<p>Application of Methods and Tools: Apply effectively the appropriate big data analytic methods and tools used to analyze big data and facilitate informed decision making (such as Rstudio, Hadoop and MapReduce).</p>	Skills: Communication, ICT and Numeracy Skills [Level 9] Knowledge: Practical Understanding [Level 9] Competence: Context [Level 8]

C. Critical-Thinking Skills		NQF Descriptor/ Level
C1	Analytic skills: Analyze sets of big data critically to generate useful insights and inform decision making.	Generic cognitive skills [Level 9]
C2	Synthetic: Identify technological requirements, architecture, and infrastructure needed to build up big data systems and address risk related issues.	Generic cognitive skills [Level 9]
C3	Creative Thinking and innovation: Identify new areas and ways to apply and communicate results of big data analytics lifecycle creatively to support insights and address emerging business challenges.	Generic cognitive skills [Level 9]

D. General and Transferable Skills (other skills relevant to employability and personal development)		NQF Descriptor/ Level
D1	Communication: Express and convey ideas rigorously and effectively, in written and oral form, to a diverse range of stakeholders using visualization techniques and tools.	Communication, ICT and Numeracy Skills [Level 9]

D2	Teamwork and Leadership: Work effectively as a member/leader of a team in order to complete a big data analytics research project.	Competence: Autonomy & Responsibility [Level 9] Competence: Learning to learn [Level 9]
D3	Organizational and Developmental Skills: Engage in life-long learning and continuing self-development to enhance and practice professional and organizational skills in order to effectively prioritize, plan, manage and allocate appropriate time to implement tasks in big data projects and assignments.	Competence: Learning to learn [Level 9]
D4	Ethics and Social Responsibility: Understand and demonstrate awareness of the benefits that big data analytics bring to organizations and the ethical and privacy risks they might pose.	Competence- Insight [Level 8]

Course Structure (Outline)

Course Structure (Outline)						
Week	Hours		ILOs	Topics	Teaching Method	Assessment Method
	Lec.	Lab				
1	3	-	A1	Introduction and overview: Structured vs. unstructured data, relational databases, big data concepts, characteristics and importance, data sciences and analytics.	Lecture	Class Participation (A1)
2	2	-	A1,B1,B2,C1,C2	Examples of big data applications and solutions for business problems: strategic planning, platform	Lecture/ case study/ discussion/ group debate	Class Participation (A1, C2)
2	1	-	A1,A3,C1,C2,D4	Examples of big data applications and solutions for business problems: security and privacy concerns, risk management.	Lecture/ case study/ discussion/ group debate	Class Participation (A1,C1,D4)
3	2	-	A1,B1,B2,C2, D3	Data Analytics Lifecycle: Discovery, Data Preparation, Model Planning, Model Building,	Lecture	Research Assignment 1 (A1,B1,B2,C2, D3) / Class

						Participation (A1)
3	1	-	A1,B1,B2,C1,D1	Data Analytics Lifecycle: Results Communication, Operationalization	Lecture	Class Participation (A1,D1)
4	2	-	A1, B2,B3,D1, D3	Basic Data Analytic Methods: Introduction to RStudio	Lecture/ Lab session demo	Class Participation (A1)
4	1	-	A1,B2, B3,C1, D1, D3	Basic Data Analytic Methods: analyzing and exploring data by R	Lecture/ Lab session demo	Project 1 (B2,B3,C1, D3) /Class Participation (A1,D1)
5	1	-	A1,B2, B3,C1, D1, D3	Basic Data Analytic Methods: analyzing and exploring data by R	Lecture/ Lab session demo	Class Participation (A1,D1)
5	2	-	A1, B2,B3,C1,C2,C3,	Basic Data Analytic Methods: statistics for model building and evaluation	Lecture	Class Participation (A1)
6	3	-	A1,B1,C1,C2, D1	Advanced Data Analytics Theory And Methods: Supervised & unsupervised learning	Lecture/use cases	Class Participation (A1)
7	2	-	A1,B1,B2,C1,C2,C3	Advanced Data Analytics Theory And Methods: clustering, association rules	Lecture/use cases	Research Assignment 2 (A1, B1, B2, C1, C2, C3)/ Class Participation (A1, C1)
7	1	-	A1,B1,B2,C1,C2,C3	Advanced Data Analytics Theory And Methods: regression, classification	Lecture/use cases	Class Participation (A1, C1)
8	2	-	A1,B1,B2,C1,C3	Advanced Data Analytics Theory And Methods: decision trees, time series analysis, text analysis	Lecture/use cases	Midterm Exam (A1,B1,B2,C1,C3) / Class

						Participation (A1, C1)
8	1	-	A1, B2, C2,C3	Advanced Data Analytics Theory And Methods: networks and graphical models, Dimension reduction	Lecture/use cases	Class Participation (A1)
9	3	-	A1,B1,B2,C1,C2	Advanced Technologies and Tools: Data storage and analytics for unstructured data	Lecture	Class Participation (A1,C1)
10	3	-	B2,B3,C1,C2,D1,D3	Advanced Technologies and Tools: MapReduce and Hadoop, Hadoop Ecosystem	Lecture/ Lab session demo	Class Participation (A1, D1)
11	3	-	A1,B2,B3,C1,C2	Tools for SQL-like access to unstructured data: Pig and Hive	Lecture/ Lab session demo	Project 2 (B2,B3,C1, C2) / Class Participation (A1)
12	3	-	A1,B2,B3,C1,C2, D3	NoSQL storage solutions: HBase, Cassandra, and Oracle NoSQL,	Lecture/ Lab session demo	Class Participation (A1)
13	3	-	A1,A3,B1,B2,C3,D1,D4	Operationalization of data Analytics Project: Creating the final deliverables, data visualization techniques, privacy, ethics, governance	Lecture	Class Participation (A3,D4)
14	3	-	A1, A2,A3, B1, C3,D1,D4	Research & Trends: Cloud & big data, SaaS business intelligence, globally distributed storage systems, Google data center infrastructure, data visualization, automation, semantics, predictive analytics, social media analytics	Lecture/ discussion/ group debate	Class Participation (A1, D4)
15	3	-	A1,A2, A3, B1,B2,B3, C1,C2,C3, D1,D2, D3, D4	Team research project report and presentations by students	Students Project Presentations	Evaluation of Research Project Outcomes and Presentations (A1,A2, A3, B1,B2,B3, C1,C2,C3, D1,D2, D3, D4)

16	2	-				Final Exam (A1, A3, B1,B2,B3, C1,C2, D4)
----	---	---	--	--	--	--

Teaching Materials:

Textbook(s):	<p>1- Nataraj Dasgupta, <i>Practical Big Data Analytics</i>, Packt Publishing, 2018.</p> <p>2- Cory Lesmeister. <i>Matering Machine Learning with R – Second Edition</i>. Packt Publishing, 2017.</p>
Handout(s):	Power point slides, http://www.ahlia.edu.bh/moodle .
Reference(s):	<ol style="list-style-type: none"> 1. Hadley Wickham, Garrett Golemund, <i>R for Data Science: Import, Tidy, Transform, Visualize, and Model Data</i>, Shroff/O'Reilly, 2017. 1. Sunila Gollapudi, <i>Practical Machine Learning</i>, Packt Publishing, 2016. 2. M. Minelli, M. Chambers and A. Dhiraj, <i>Fundamentals of Deep Learning: Designing next generation Machine Intelligence Algorithms</i>, Shroff/O'Reilly, 2017. 3. Sam R. Alapati, <i>Expert Hadoop Administration: Managing, Tuning, and Securing Spark, YARN, HDFS</i>, First Edition, Pearson Education, 2017. 4. Gurmukh Singh, <i>Hadoop 2.x Administration Cookbook</i>, Packt Publishing, 2017. 5. Dipayan Dev, <i>Deep Learning with Hadoop</i>, Packt Publishing, 2017. 6. Dr. PKS Prakash, <i>R Deep Learning Cookbook</i>, Packt Publishing, 2017. 7. Tom M. Mitchell, <i>Machine Learning</i>, First edition, McGraw Hill Education, 2017. 8. http://www.rstudio.com/ 9. https://mran.microsoft.com/open 10. http://tdwi.org/ 11. More references are available in the course website in Moodle.

Assessment

Type of Assessment ¹	Description ²	ILOs ³	Weighting
Research Assignment 1	The assignment consists of 6-8 essay, problem-solving and research based theoretical questions regarding topics in big data analytics covered during the first 3 weeks. The purpose of the assignment is to assess students individually where they have to demonstrate	A1, B1, B2, C2, D3	7%

	<p>their extensive and detailed knowledge and critical understanding of key concepts, theories, methods, tools and techniques of big data analytics. The assignment will also assess students' skills to solve and develop solutions for business-related problems by applying the big data analytics lifecycle. Soft copy submission is required by the end of the 4th week through the course page in Moodle where answers will be checked by Turnitin against plagiarism.</p>		
Research Assignment 2	<p>Similar to assignment 1 but topics will cover up to week 7. Soft copy submission is required by the end of the 8th week through the course page in Moodle where answers will be checked by Turnitin against plagiarism.</p>	A1, B1, B2, C1, C2, C3	7%
Project 1	<p>The practical project will assess students' skills in utilizing the R language and RStudio to analyze and explore big data. The students will be given examples of big data sets and will be asked to analyze them using R language and RStudio. Soft copy of the output should be submitted electronically by the end of week 5.</p>	B2,B3,C1, D3	7%
Project 2	<p>Students will be given a set of practical tasks to be implemented on real examples of big data using MapReduce, Hadoop, Pig and Hive. The purpose of the project is to assess students' skills (individually) in using the advanced big data analytics techniques and methods learned in class. Soft copy of the output should be submitted electronically by the end of week 12.</p>	B2,B3,C1, C2	7%
Midterm Exam	<p>The midterm exam will be an in-class 1-hour exam that will consists of short-answer, essay, and problem solving questions and cover the topics studied in the first 8 weeks.</p>	A1,B1,B2,C 1,C3	20%
Team Research Project Report and Presentation	<p>Starting from weak 4, the class will be divided into teams of 4-5 students and each will be asked to study a research problem in big data analytics utilizing the skills and tools learned in class. Teams are required to plan and execute the research project that may involve the application of advanced and specialized big data techniques, skills and tools to develop solutions for business-related problems by applying the big data analytics lifecycle. In the final week, each team will have to submit their research report (worth 10%) explaining the research problem, research methods used, analysis and the conclusion highlighting the research findings and results. The report must explain precisely the work accomplished by each student. Each team will be required to make a presentation (worth 6%) summarizing the research conducted and its findings. Each team member has to participate in the presentation. Team-based work will be examined and evaluated as a whole as well as the individual work of each student. Team members will be tested individually during the presentation by peers and the instructor.</p>	A1,A2, A3, B1,B2,B3, C1,C2,C3, D1,D2, D3, D4	12%

Final Exam	The final exam is a 2-hour closed book and notes comprehensive exam. It will consist of short-answer, essay and problem-solving questions in big data analytics.	A1, A3, B1,B2,B3, C1,C2,C3, D4	40%
Overall			100%

Admissions	
Pre-requisites	ITCS 511
Minimum number of students	4
Maximum number of students	20