MODULE DESCRIPTION FORM

نموذج وصف المادة الدراسية

| **Module Information**  **معلومات المادة الدراسية** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Module Title** | Big Data Analytics | | | | **Module Delivery** | | |
| **Module Type** | Elective | | | | * **☒ Theory** * **☐ Lecture** * **☒ Lab** * **☐ Tutorial** * **☐ Practical** * **☐ Seminar** | | |
| **Module Code** | BMI421 | | | |
| **ECTS Credits** | 5.00 | | | |
| **SWL (hr/sem)** | 125 | | | |
| **Module Level** | | 4 | **Semester of Delivery** | | | | 8 |
| **Administering Department** | | BID | **College** | BMIC | | | |
| **Module Leader** | Ahmed Oday Ezzat | | **e-mail** | Ahmed.oday@uoitc.edu.iq | | | |
| **Module Leader’s Acad. Title** | | Asst.Lect | **Module Leader’s Qualification** | | | | Master |
| **Module Tutor** | \ | | **e-mail** | \ | | | |
| **Peer Reviewer Name** | | omar A.M | **e-mail** | omara.m@uoitc.edu.iq | | | |
| **Scientific Committee Approval Date** | | 18/6/2023 | **Version Number** | | | \ | |

| **Relation with other Modules**  **العلاقة مع المواد الدراسية الأخرى** | | | |
| --- | --- | --- | --- |
| **Prerequisite module** | Cloud computing / BMI411 | **Semester** | 7 |
| **Co-requisites module** | None | **Semester** |  |

| **Module Aims, Learning Outcomes and Indicative Contents**  **أهداف المادة الدراسية ونتائج التعلم والمحتويات الإرشادية** | |
| --- | --- |
| **Module Objectives**  **أهداف المادة الدراسية** | 1. To understand the Big Data Platform and its Use cases. 2. To provide an overview of Apache Hadoop. 3. To provide HDFS Concepts and Interfacing with HDFS. 4. To understand Map Reduce Jobs. 5. To provide hands on Hodoop EcoSystem. 6. To apply analytics on Structured, Unstructured Data. |
| **Module Learning Outcomes**  **مخرجات التعلم للمادة الدراسية** | The students will be able to:   1. Identify Big Data and its Business Implications. 2. List the components of Hadoop and Hadoop Ecosystem. 3. Access and Process Data on Distributed File System. 4. Manage Job Execution in Hadoop Environment. 5. Develop Big Data Solutions using Hadoop EcoSystem. 6. Analyze Infosphere BigInsights Big Data Recommendations. |
| **Indicative Contents**  **المحتويات الإرشادية** | Indicative content includes the following.  Part A – Why Big Data  Data and bigdata, Details and different type of data and big data with characteristic data  Big Data is the hot new buzzword in IT circles. The proliferation of digital technologies with digital storage and recording media has created massive amounts of diverse data, which can be used for marketing and many other purposes  Part B - Analogue Electronics  Big Data refers to massive and often unstructured data, on which the processing capabilities of traditional data management tools result to be inadequate.  Big Data can take up terabytes and petabytes of storage space in diverse formats including text, video, sound, images, and more. |

| **Learning and Teaching Strategies**  **استراتيجيات التعلم والتعليم** | |
| --- | --- |
| **Strategies** | Learn the fundamentals: Start by developing a strong foundation in the fundamental concepts of big data. This includes understanding data structures, data modeling, data management, and data processing techniques. Familiarize yourself with concepts such as Hadoop, MapReduce, and distributed computing |

| **Student Workload (SWL)**  **الحمل الدراسي للطالب محسوب لـ ١٥ اسبوعا** | | | |
| --- | --- | --- | --- |
| **Structured SWL (h/sem)**  **الحمل الدراسي المنتظم للطالب خلال الفصل** | 64 | **Structured SWL (h/w)**  **الحمل الدراسي المنتظم للطالب أسبوعيا** | 4 |
| **Unstructured SWL (h/sem)**  **الحمل الدراسي غير المنتظم للطالب خلال الفصل** | 61 | **Unstructured SWL (h/w)**  **الحمل الدراسي غير المنتظم للطالب أسبوعيا** | 4 |
| **Total SWL (h/sem)**  **الحمل الدراسي الكلي للطالب خلال الفصل** | **125** | | |

| **Module Evaluation**  **تقييم المادة الدراسية** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **As** | | **Time/Number** | **Weight (Marks)** | **Week Due** | **Relevant Learning Outcome** |
| **Formative assessment** | **Quizzes** | 4 | 10% (10) | 2,4,6 and 10 | #1, 2, 3 |
| **Assignments** | 2 | 10% (10) | 5 and 9 | #3, 4, 5 |
| **Projects / Lab.** | 1 | 20% (20) | Continuous | All |
| **Summative assessment** | **Midterm Exam** | 2hr | 10% (10) | 7 | LO #1, 3, 5 |
| **Final Exam** | 3hr | 50% (50) | 16 | All |
| **Total assessment** | | | 100% (100 Marks) |  |  |

| **Delivery Plan (Weekly Syllabus)**  **المنهاج الاسبوعي النظري** | |
| --- | --- |
| **Week** | **Material Covered** |
| **Week 1** | Big data introduction |
| **Week 2** | Characteristics of Big data |
| **Week 3** | Hdfs(hadoop distributed file system) |
| **Week 4** | Predictive Analytics on Big Data |
| **Week 5** | Getting Value out of Big Data |
| **Week 6** | First exam |
| **Week 7** | Introduction to Spark & Architecture |
| **Week 8** | Spark transformation & Action |
| **Week 9** | What is machine learning |
| **Week 10** | Linear Regression |
| **Week 11** | Machine Learning Algorithm K-means using Map Reduce for Big Data Analytics |
| **Week 12** | Parallel K-means using Map Reduce on Big Data Cluster Analysis |
| **Week 13** | Case study session |
| **Week 14** | Case study session |
| **Week 15** | Second exam |

| **Delivery Plan (Weekly Lab. Syllabus)**  **المنهاج الاسبوعي للمختبر** | |
| --- | --- |
| **Week** | **Material Covered** |
| **Week 1-2** | Introduction of Hadoop architecture |
| **Week 3-4** | Install Hadoop |
| **Week 5-6** | HDFS architecture |
| **Week 7-8** | Distributed Processing MapReduce Framework |
| **Week 9-10** | Decision Trees for Big Data Analytics |
| **Week 11-12** | MapReduce Advanced Concepts |
| **Week 13-14** | Data Ingestion into Big Data Systems and ETL |
| **Week 15** | Review and Discussions |

| **Learning and Teaching Resources**  **مصادر التعلم والتدريس** | | |
| --- | --- | --- |
|  | **Text** | **Available in the Library?** |
| **Required Texts** | * Matthew J. Salganik. (2017). Bit by Bit: Social Research in the Digital Age. Princeton University Press. * Cathy O’Neil. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Penguin Books. * Rob Kitchin. (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. SAGE Publications | No |

| **Grading Scheme**  **مخطط الدرجات** | | | | |
| --- | --- | --- | --- | --- |
| **Group** | **Grade** | **التقدير** | **Marks %** | **Definition** |
| **Success Group**  **(50 - 100)** | **A -** Excellent | **امتياز** | 90 - 100 | Outstanding Performance |
| **B -** Very Good | **جيد جدا** | 80 - 89 | Above average with some errors |
| **C -** Good | **جيد** | 70 - 79 | Sound work with notable errors |
| **D -** Satisfactory | **متوسط** | 60 - 69 | Fair but with major shortcomings |
| **E -** Sufficient | **مقبول** | 50 - 59 | Work meets minimum criteria |
| **Fail Group**  **(0 – 49)** | **FX –** Fail | **راسب (قيد المعالجة)** | (45-49) | More work required but credit awarded |
| **F –** Fail | **راسب** | (0-44) | Considerable amount of work required |
|  |  |  |  |  |
| **Note:** Marks Decimal places above or below 0.5 will be rounded to the higher or lower full mark (for example a mark of 54.5 will be rounded to 55, whereas a mark of 54.4 will be rounded to 54. The University has a policy NOT to condone "near-pass fails" so the only adjustment to marks awarded by the original marker(s) will be the automatic rounding outlined above. | | | | |