

| | |
|--|---|
| Short Title: | Programming for Big Data APPROVED |
| Full Title: | Programming for Big Data |
| Module Code: | ADSA H6018 |
| ECTS credits: | 10 |
| NFQ Level: | 9 |
| Module Delivered in | 2 programme(s) |
| Module Contributor: | Simon McLoughlin |
| Module Description: | <p>Students taking this module will acquire the computer programming skills necessary to analyse and manipulate big data. Big data in this context refers to datasets that are too large to be handled by the software tools commonly used to analyse and manipulate data within a tolerable elapsed time. The algorithms and challenges for processing large datasets form a core part of this course, such that the student will be able to select the appropriate algorithms, tools or methods for big data problems in addition to being able to implement and evaluate solutions using a variety of programming techniques and tools. Students are not expected to have advanced programming skills in order to take the module, but will need to have fundamental knowledge and skills in computer programming.</p> |
| Learning Outcomes: | |
| <i>On successful completion of this module the learner will be able to</i> | |
| <ol style="list-style-type: none"> 1. Clearly describe the characteristics of big data, and contrast the requirements for processing big data with conventional data. 2. Identify and illustrate the challenges of programming for big data, and evaluate contrasting methods for addressing these challenges. 3. Demonstrate a detailed understanding of the state of the art in Big Data algorithms and techniques. 4. Select and evaluate the appropriate development tools for various big data programming problems. 5. Demonstrate a detailed understanding of state of the art distributed programming paradigms for both data storage and data analysis, and select the appropriate method for a given context. 6. Implement solutions to various big data programming problems using a range of state of the art tools and techniques, and evaluate the effectiveness of these solutions. 7. Present an informed view of the changing big data landscape and how programming for big data may change in the future, based on current literature and standards. | |

Module Content & Assessment

| |
|---|
| Indicative Content |
| Introduction to Big Data Programming Study of Big Datasets. Programming for big data versus traditional data access programming. Big Data Visualisation. |
| Programming languages for big data analysis Learning the constructs of a programming language suitable for Big Data Analysis, e.g. Python, R, noSQL. |
| Big Data Algorithms and Data Structures MapReduce, Page Rank extensions, Market Basket Models etc |
| Real-time Stream processing Hashing Techniques, Online Algorithms, Competitive Ration. |
| Tools for Big Data analysis (e.g. Hadoop, Mahoot, Pig, noSQL). n/a |
| Cloud Services for Big Data Identification and Evaluation of different cloud services available for Big Data Analysis, e.g. IBM Bluemix, Big Insights, InfoSphere Stream, Amazon EC2 etc. |
| Future trends in Big Data n/a |

| | |
|--|----------|
| Indicative Assessment Breakdown | % |
| Course Work Assessment % | 100.00% |

| Course Work Assessment % | | | | |
|---------------------------------|--|--------------------------|-------------------|------------------------|
| <i>Assessment Type</i> | <i>Assessment Description</i> | <i>Outcome addressed</i> | <i>% of total</i> | <i>Assessment Date</i> |
| Lab work | Regular exercises in Big Data Programming. | 2,3,4,5,6 | 40.00 | n/a |
| Project | A major project based assignment where students implement a Big Data solution and execute it on an appropriate platform (e.g. a cloud service, virtual cluster etc.) | 2,3,4,5,6,7 | 30.00 | n/a |
| Written Report | Technical Review(s) in the state of the art of Big Data Analytics | 1,2,3,7 | 30.00 | n/a |

| |
|----------------------------|
| No Final Exam Assessment % |
|----------------------------|

| |
|--|
| Indicative Reassessment Requirement |
| Coursework Only <i>This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.</i> |
| Reassessment Description Major Project in Big Data Technical Reports on State of the Art Number of smaller lab based exercises |

ITB reserves the right to alter the nature and timings of assessment

Indicative Module Workload & Resources

Indicative Workload: Part Time

| Frequency | Indicative Average Weekly Learner Workload |
|------------|--|
| Every Week | 30.00 |
| Every Week | 30.00 |
| Every Week | 40.00 |

Resources

Recommended Book Resources

Anand Rajaraman, Jeffrey David Ullman, *Mining of Massive Datasets*, Cambridge University Press [ISBN: 1107015359]

This module does not have any article/paper resources

Other Resources

website: *Big Data University* (<http://bigdatauniversity.com/>).

website: *UCI dataset repository*
<https://archive.ics.uci.edu/ml/datasets.html>

Journal: *Big Data Research Journal (ELSEVIER)*

Journal: *Journal of Big Data (Springer)*

Module Delivered in

| Programme Code | Programme | Semester | Delivery |
|----------------|---|----------|----------|
| BN_EMIOT_R | Master of Engineering in Internet of Things Technologies [BN535R 60 credits taught with a 30 credit research project] | 2 | Elective |
| BN_KADSA_R | Master of Science in Computing in Applied Data Science & Analytics | 3 | Elective |