

## H9SSP: Scalable Systems Programming

|                                                                              |                                                                                                          |
|------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|
| <b>Module Code:</b>                                                          | H9SSP                                                                                                    |
| <b>Long Title</b>                                                            | Scalable Systems Programming <b>APPROVED</b>                                                             |
| <b>Title</b>                                                                 | Scalable Systems Programming                                                                             |
| <b>Module Level:</b>                                                         | LEVEL 9                                                                                                  |
| <b>EQF Level:</b>                                                            | 7                                                                                                        |
| <b>EHEA Level:</b>                                                           | Second Cycle                                                                                             |
| <b>Credits:</b>                                                              | 5                                                                                                        |
| <b>Module Coordinator:</b>                                                   | Horacio Gonzalez-Velez                                                                                   |
| <b>Module Author:</b>                                                        | Margarete Silva                                                                                          |
| <b>Departments:</b>                                                          | School of Computing                                                                                      |
| <b>Specifications of the qualifications and experience required of staff</b> | MSc and/or PhD degree in computer science or cognate discipline. May have industry experience also.      |
| <b>Learning Outcomes</b>                                                     |                                                                                                          |
| <i>On successful completion of this module the learner will be able to:</i>  |                                                                                                          |
| <b>#</b>                                                                     | <b>Learning Outcome Description</b>                                                                      |
| LO1                                                                          | Demonstrate in-depth knowledge of parallel algorithms on large amounts of data                           |
| LO2                                                                          | Identify and categorise search techniques including similarity search and search engine technologies.    |
| LO3                                                                          | Critically compare and contrast different data-stream processing and specialised algorithms.             |
| LO4                                                                          | Critically analyse mining and clustering algorithms on large multi-dimensional datasets.                 |
| LO5                                                                          | Develop and implement efficient programming solutions for problems relating to processing data at scale. |
| <b>Dependencies</b>                                                          |                                                                                                          |
| <b>Module Recommendations</b>                                                |                                                                                                          |
| No recommendations listed                                                    |                                                                                                          |
| <b>Co-requisite Modules</b>                                                  |                                                                                                          |
| No Co-requisite modules listed                                               |                                                                                                          |
| <b>Entry requirements</b>                                                    | A level 8 degree or its equivalent in any discipline                                                     |

# H9SSP: Scalable Systems Programming

| Module Content & Assessment                                                                                                                                                                                                                                                                                 |                       |                           |            |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|---------------------------|------------|
| <b>Indicative Content</b>                                                                                                                                                                                                                                                                                   |                       |                           |            |
| <b>MapReduce Extensions</b><br>Recursive and workflow systems for MapReduce. Resilient data sets.                                                                                                                                                                                                           |                       |                           |            |
| <b>MapReduce Cost Models</b><br>Complexity and cost models for MapReduce with emphasis on communication costs and task networks                                                                                                                                                                             |                       |                           |            |
| <b>Near Neighbour search and Shingling</b><br>Collaborative filtering and similarity sets. Document shingling and sub-strings.                                                                                                                                                                              |                       |                           |            |
| <b>Hashing</b><br>Locality-sensitive hashing and distance measures. Additional methods for higher degrees of similarity.                                                                                                                                                                                    |                       |                           |            |
| <b>Stream Data Model</b><br>Stream sources, stream queries, and processing. Sampling data                                                                                                                                                                                                                   |                       |                           |            |
| <b>Streams Operations</b><br>Filtering, counting, combining and estimating.                                                                                                                                                                                                                                 |                       |                           |            |
| <b>Stream Processing</b><br>Building complex pipelines and models                                                                                                                                                                                                                                           |                       |                           |            |
| <b>Link Analysis</b><br>PageRank algorithm in its application to search engines. Efficient computation of PageRank. Link Spam. Hubs and authorities.                                                                                                                                                        |                       |                           |            |
| <b>Frequent itemsets</b><br>Market-Basket Model, many-to-many relationships. Association rules.                                                                                                                                                                                                             |                       |                           |            |
| <b>A-Priori / Limited Pass Algorithms</b><br>Determine stages, sets and items under different monotonicity conditions.                                                                                                                                                                                      |                       |                           |            |
| <b>Clusters for Streams and Parallelism</b><br>Bucket initialisation and merging. Parallel clustering.                                                                                                                                                                                                      |                       |                           |            |
| <b>Using Scalable Services</b><br>Deploying concurrent stream processing and batch processing pipelines                                                                                                                                                                                                     |                       |                           |            |
| <b>Assessment Breakdown</b>                                                                                                                                                                                                                                                                                 |                       |                           | %          |
| Coursework                                                                                                                                                                                                                                                                                                  |                       |                           | 100.00%    |
| Assessments                                                                                                                                                                                                                                                                                                 |                       |                           |            |
| Full Time                                                                                                                                                                                                                                                                                                   |                       |                           |            |
| <b>Coursework</b>                                                                                                                                                                                                                                                                                           |                       |                           |            |
| <b>Assessment Type:</b>                                                                                                                                                                                                                                                                                     | Formative Assessment  | <b>% of total:</b>        | Non-Marked |
| <b>Assessment Date:</b>                                                                                                                                                                                                                                                                                     | n/a                   | <b>Outcome addressed:</b> | 1,2,3,4,5  |
| <b>Non-Marked:</b>                                                                                                                                                                                                                                                                                          | Yes                   |                           |            |
| <b>Assessment Description:</b><br>Formative assessment will be provided on the in-class individual or group activities. Feedback will be provided in written or oral format, or on-line through Moodle. In addition, in class discussions will be undertaken as part of the practical approach to learning. |                       |                           |            |
| <b>Assessment Type:</b>                                                                                                                                                                                                                                                                                     | Continuous Assessment | <b>% of total:</b>        | 100        |
| <b>Assessment Date:</b>                                                                                                                                                                                                                                                                                     | n/a                   | <b>Outcome addressed:</b> | 1,2,3,4,5  |
| <b>Non-Marked:</b>                                                                                                                                                                                                                                                                                          | No                    |                           |            |
| <b>Assessment Description:</b><br>This practical assessment will evaluate the learners' knowledge and understanding of Scalable Systems Programming, possibly in the context of mining and/or clustering algorithms.                                                                                        |                       |                           |            |
| No End of Module Assessment                                                                                                                                                                                                                                                                                 |                       |                           |            |
| No Workplace Assessment                                                                                                                                                                                                                                                                                     |                       |                           |            |
| Reassessment Requirement                                                                                                                                                                                                                                                                                    |                       |                           |            |
| <b>Coursework Only</b><br><i>This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.</i>                                                                                                                                                          |                       |                           |            |
| <b>Reassessment Description</b><br>Reassessment of this module will be via project                                                                                                                                                                                                                          |                       |                           |            |

## H9SSP: Scalable Systems Programming

| Module Workload                      |                                    |       |            |                                 |
|--------------------------------------|------------------------------------|-------|------------|---------------------------------|
| Module Target Workload Hours 0 Hours |                                    |       |            |                                 |
| Workload: Full Time                  |                                    |       |            |                                 |
| Workload Type                        | Workload Description               | Hours | Frequency  | Average Weekly Learner Workload |
| Lecture                              | Classroom & Demonstrations (hours) | 24    | Every Week | 24.00                           |
| Tutorial                             | Other hours (Practical/Tutorial)   | 24    | Every Week | 24.00                           |
| Independent Learning                 | Independent learning (hours)       | 77    | Every Week | 77.00                           |
| Total Weekly Contact Hours           |                                    |       |            | 48.00                           |

| Module Resources                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|
| <i>Recommended Book Resources</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |  |
| <p>Jure Leskovec, Anand Rajaraman, Jeffrey David Ullman. (2014), Mining of Massive Datasets, Cambridge University Press, p.476, [ISBN: 1107077230].</p> <p>Martin Kleppmann. (2017), Designing Data-intensive Applications, O'Reilly &amp; Associates Incorporated, p.590, [ISBN: 1449373321].</p>                                                                                                                                                                                                                                                                                                                                               |  |
| <i>Supplementary Book Resources</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |  |
| <p>Andrew Kelleher, Adam Kelleher. (2018), Machine Learning in Production, [ISBN: 9780134116556].</p> <p>K. Hwang. (2017), Cloud and Cognitive Computing: A Machine Learning Approach, MIT Press, [ISBN: 10026203641X].</p> <p>B. Chambers, M. Zaharia. (2018), Spark - The Definitive Guide, 1st. O'Reilly Media, [ISBN: 101491912219].</p> <p>Tom White. Hadoop: the Definitive Guide ; Storage and Analysis at Internet Scale, [ISBN: 1491901632].</p>                                                                                                                                                                                        |  |
| <i>Recommended Article/Paper Resources</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |  |
| <p>B. Veloso, F. Leal, H. González-Vélez, B. Malheiro, J-C. Burguillo. (2018), Scalable data analytics using crowdsourced repositories and streams, Journal of Parallel and Distributed Computing, 122, p.1-10.</p> <p>J. Eckroth. (2018), A course on big data analytics, Journal of Parallel and Distributed Computing, 118, p.166.</p> <p>J. Kołodziej, H. González-Vélez, H.D. Karatza. (2017), High-performance modelling and simulation for big data applications, Simulation Modelling Practice and Theory, 76, p.1-2.</p> <p>J. Dean, S. Ghemawat. (2010), MapReduce: a flexible data processing tool., Commun. ACM, 53(1), p.72-77.</p> |  |
| <i>This module does not have any other resources</i>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             |  |
| Discussion Note:                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                 |  |