H9SCP: Scalable Cloud Programming

Module Code:	H9SCP
Long Title	Scalable Cloud Programming APPROVED
Title	Scalable Cloud Programming
Module Level:	LEVEL 9
EQF Level:	7
EHEA Level:	Second Cycle
Credits:	10
Module Coordinator:	Horacio Gonzalez-Velez
Module Author:	Noel Cosgrave
Departments:	
Specifications of the qualifications and experience required of staff	MSc and/or PhD degree in computer science or cognate discipline. Experience lecturing in the field. May have industry experience also.

Learning Outcomes

On successful completion of this module the learner will be able to:

#	Learning Outcome Description
LO1	Identify and critically evaluate functional and non-functional characteristics of parallel workloads on cloud platforms
LO2	Analyse sequential programs to identify suitable candidates for parallelisation.
LO3	Demonstrate competence in writing parallel programs using scalable algorithms and techniques
LO4	Recognise and describe techniques and tools to improve the productivity of parallel programming on emerging computing architectures
LO5	Identify and critically evaluate system-specific levels of parallelism and co-scheduling of computation for scalability and resilience.

Dependencies

Module Recommendations

No recommendations listed

Co-requisite Modules

No Co-requisite modules listed

Entry requirements	Internal to the programme

H9SCP: Scalable Cloud Programming

Module Content & Assessment

Indicative Content
Parallel Programming Preliminaries Sequential Algorithms vs. Parallel algorithms. Parallelism vs Concurrency. Process Management. Multitasking.
Parallel and Distributed Software Threads. Coordinating Process and Threads. Shared-Memory. Distributed Memory. Programming Hybrid Systems.
Scaling Deployments Paradigms of Parallel Computing in the Cloud. SPMD and HPC-style parallelism. Many-Task Parallelism.
Parallel Software Patterns Data- vs. Process Parallelism. Data patterns: Map, reduce, scan, gather. Collectives. Task Farms and Pipelines.
MapReduce MapReduce and Graph Data-flows. Recursive and workflow systems for MapReduce.
MapReduce Cost Models Complexity and cost models for MapReduce with emphasis on communication costs and task networks.
Multi-stage and data-flow computing Resilient Distributed Data Sets (RDDs). RDDs vs DAG Tasks.
Streaming Data Model Stream sources, stream queries, and processing. Sampling data.
Stream Operations Filtering, counting, combining and estimating.
Stream Processing Building complex pipelines and models
Cloud Performance Metrics and Benchmarks. Autoscaling, Scale-Out, Scale-up and Mixed Scaling. Scaling Strategies.
Using Scalable Services Deploying concurrent stream processing and batch processing pipelines

Assessment Breakdown	%
Coursework	50.00%
End of Module Assessment	50.00%

Assessments

Full Time

Coursework

Assessment Type:	Project	**% of total:**	50
Assessment Date:	n/a	**Outcome addressed:**	3,4,5
Non-Marked:	No		

Assessment Description:
Develop a complex scalable cloud computing solution, which should be informed by a review of recent work in the domain, and should be submitted in the form of a conference-style report The working solution will be demonstrated to the lecturer, either by means of a project video or in-class presentation. Marked elements include the methodology, implementation, clarity of presentation and depth of understanding of the work carried out and its broader implications.

Assessment Type:	Written (0080)	**% of total:**	50
Assessment Date:	n/a	**Outcome addressed:**	1,2,3
Non-Marked:	No		

Assessment Description:
The test will assess learners' knowledge and understanding of data and computing architectures, programming models, and storage concepts. A sample question, marking scheme, and solution, is provided in Appendices.

No End of Module Assessment

No Workplace Assessment

Reassessment Requirement

Repeat examination
Reassessment of this module will consist of a repeat examination. It is possible that there will also be a requirement to be reassessed in a coursework element.

Reassessment Description
Reassessment of this module will be via proctored examination.

H9SCP: Scalable Cloud Programming

Module Workload

Module Target Workload Hours 0 Hours

Workload: Full Time

Workload Type	*Workload Description*	*Hours*	*Frequency*	*Average Weekly Learner Workload*
Lecture	No Description	36	Per Semester	3.00
Tutorial	No Description	24	Per Semester	2.00
Independent Learning	No Description	190	Per Semester	15.83
			Total Weekly Contact Hours	5.00

Module Resources

Recommended Book Resources

Ian Foster,Dennis B. Gannon. (2017), Cloud Computing for Science and Engineering, MIT Press, p.392, [ISBN: 978-0-262-03724-2].

Supplementary Book Resources

Peter Pacheco. (2019), An Introduction to Parallel Programming, 2nd Edition. Morgan Kaufmann, Amsterdam, [ISBN: 0128046058].

Kai Hwang. (2017), Cloud Computing for Machine Learning and Cognitive Applications, MIT Press, Cambridge, MA., [ISBN: 026203641X].

K.C. Wang. (2018), Systems Programming in Unix/Linux, Springer, [ISBN: 978-3-319-92428-1].

H Karau et al.. (2015), Learning Spark, 1st edition. O'Reilly Media, [ISBN: 1449358624].

Tom White. (2015), Hadoop: The Definitive Guide, 4th Edition. O'Reilly Media, [ISBN: 1449311520].

Maurice Herlihy,Nir Shavit. (2012), The Art of Multiprocessor Programming, Revised Edition. Morgan Kaufmann, Amsterdam, [ISBN: 0123973376].

William Gropp,Ewing Lusk,Anthony Skjellum. (2015), Using MPI: Portable Parallel Programming with the Message-Passing Interface (Scientific and Engineering Computation), 3rd Edition. MIT Press, Cambridge, MA, [ISBN: 0262527391].

Ruud van der Pas,Eric Stotzer,Christian Terboven. (2017), Using OpenMP: The Next Step: Affinity, Accelerators, Tasking, and SIMD (Scientific and Engineering Computation), 1st Edition. MIT Press, Cambridge, MA, [ISBN: 0262534789].

Joanna Kołodziej (Editor), Horacio González-Vélez (Editor). (2019), High-Performance Modelling and Simulation for Big Data Applications: Selected Results of the COST Action IC1406 cHiPSet, 1. Springer, Cham, [ISBN: 3030162710].

Recommended Article/Paper Resources

R. Buyya et al.. (2019), Manifesto for Future Generation Cloud Computing: Research Directions for the Next Decade., ACM Computing Surveys, 51(5), [ISSN: 0360-0300].

J. Dean, S. Ghemawat. (2010), MapReduce: a flexible data processing tool., Communications of the ACM, 53(1), [ISSN: 0001-0782].

H. González-Vélez, M. Leyton. (2010), A survey of algorithmic skeleton frameworks: high-level structured parallel programming enablers, Software: Practice and Experience, 40(12), [ISSN: 0038-0644].

N. P. Jouppi et al.. (2017), In-datacenter performance analysis of a tensor processing unit.

This module does not have any other resources

Discussion Note:	