# **H8DIA: Data Intensive Architectures**

Module Code:		H8DIA				
Long Title		Data Intensive Architectures APPROVED				
Title		Data Intensive Architectures				
Module Level:		LEVEL 9				
EQF Level:		7				
EHEA Level:		Second Cycle				
Credits:		5				
Module Coordinator:		cio Gonzalez-Velez				
Module Author:		Margarete Silva				
Departments:		School of Computing				
Specifications of the qualifications and experience required of staff		PhD in a computer science or cognate discipline. May have industry experience also.				
Learning Outcomes						
On successful completion of this module the learner will be able to:						
#	Learning Outcome	ning Outcome Description				
LO1	Critically compare ar	nd contrast multiple distributed system models and their associated enabling technologies.				
LO2	Demonstrate in-dept	Demonstrate in-depth knowledge of different types of processing on different data-intensive computational resources.				
LO3	Identify and categori	dentify and categorise platforms and software environments for cloud and cognitive computing.				
LO4	Critically analyse the	ically analyse the features of high performance computing platforms and how they enable parallel and distributed programming paradigms.				
Dependencies	•					
Module Recommendations						
No recommendations listed						
Co-requisite Modules						
No Co-requisite modules listed						
Entry requirements		A level 8 degree or its equivalent in any discipline				

## **H8DIA: Data Intensive Architectures**

#### **Module Content & Assessment**

### Indicative Content

**Principles of Cloud Computing Systems** 

Distributed systems, service models, ecosystems

Non-functional characteristics of cloud systems

SLAs/QoS, Availability, Mobility, and Optimisation for Cloud

**Data Analytics and Cognitive Computing** 

The Big Data Industry. Data Collection, Mining and Analytics on Clouds. Neuromorphic hardware and Cognitive Computing.

**Computing Architectures** 

Multi/Many core, Clusters, Grids, and Clouds; NIST Model: Elastic provisioning, resource metering, pools, etc.

**Cloud Infrastructures and Services** 

Computation, storage and general resource deployment; Public cloud services (e.g. AWS and GAE service offerings); Machine Learning support;

Clouds for Mobile and IOT services

Mobile devices and edge computing; Mobile clouds and colocation; Mobile networks; IoT interaction frameworks

Clouds for Social Media and Mashup Services

Social media industrial applications; Social media networks and APIs; Graph analysis; Mashup architectures; Dynamic composition of services.

Structured parallel programming

Algorithmic skeletons and; Structured parallelism; scalable models; fine-grained vs. coarse-grained parallelisation

Parallel patterns for data-intensive computations

Data-enabled patterns and skeletons: map, reduce, broadcast, scan, gather scatter. MapReduce compute engine. MapReduce computations.

Data-intensive storage management

CAP Theorem; distributed file organisations: HDFS and Resilient distributed data sets;

Streams and Graphs

Structured sources; data streams; stream programming, libraries and applications; graphs; centrality and degrees; graph programming, libraries, and applications

Non Von Neumann architectures for machine learning

GPGPU; Neuromorphic hardware; TensorFlow; Cognitive services

Assessment Breakdown	%
Coursework	100.00%

#### Assessments

### **Full Time**

Coursework

**Assessment Type** 430 **Assessment Date:** n/a

% of total: Non-Marked Outcome addressed: 1.2.3.4

Non-Marked: Yes

**Assessment Description:** 

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.

100 % of total: Assessment Type: Project **Assessment Date:** n/a Outcome addressed: 1,2,3,4

Non-Marked:

**Assessment Description:** 

Produce a portfolio of studies that critically compare the data and computing architectures, programming models, and storage concepts

No End of Module Assessment

No Workplace Assessment

## Reassessment Requirement

Coursework Only

This module is reassessed solely on the basis of re-submitted coursework. There is no repeat written examination.

Reassessment Description
Reassessment of this module will be via a project

## **H8DIA: Data Intensive Architectures**

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

#### Recommended Book Resources

Kai Hwang. (2017), Cloud Computing and Cognitive Computing: A Machine Learning Approach, MIT Press, p.624, [ISBN: 026203641X].

Jan Kunigk,lan Buss,Lars George,Paul Wilkinson. (2019), Architecting Modern Data Platforms, O'Reilly Media, p.636, [ISBN: 149196927X].

#### Supplementary Book Resources

Bill Chambers, Matei Zaharia. Spark, [ISBN: 1491912219].

lan Foster, Dennis B. Gannon. (2017), Cloud Computing for Science and Engineering, MIT Press, p.392, [ISBN: 9780262037242].

Martin Kleppmann. (2017), Designing Data-intensive Applications, Oreilly & Associates Incorporated, p.590, [ISBN: 1449373321].

Michael D. McCool, Arch D. Robison, James Reinders. (2012), Structured Parallel Programming, Elsevier, p.406, [ISBN: 0124159931].

K.C. Wang. (2018), Systems Programming in Unix/Linux, Springer, p.344, [ISBN: 3319924281].

Tom White. Hadoop: the Definitive Guide; Storage and Analysis at Internet Scale, [ISBN: 1491901632].

#### Recommended Article/Paper Resources

- J. Eckroth. (2018), A course on big data analytics, Journal of Parallel and Distributed Computi, 118, p.166.
- J. Kolodziej, 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.

R. Buyya, et al. (2017), A Manifesto for Future Generation Cloud Computing: Research Directions for the Next Decade, Working Draft: Distributed, Parallel, and Cluster Computing (cs.DC), arXiv:1711.09123 [cs.DC], 43, https://arxiv.org/abs/1711.09123

- H. González-Vélez. (2010), A survey of algorithmic skeleton frameworks: high-level structured parallel programming enablers., Practice and Experience, 40(12), p.1135.
- J. Dean, S. Ghemawat. (2010), MapReduce: a flexible data processing tool., Communications of the ACM, 53(1), p.72.

This module does not have any other resources

Discussion I	Not	e:
--------------	-----	----