



Title:	NoSQL Data Architectures APPROVED
Long Title:	NoSQL Data Architectures
Module Code:	COMP7037
Duration:	1 Semester
Credits:	5
NFQ Level:	Intermediate
Field of Study:	Computer Science
Valid From:	Semester 1 - 2017/18 (September 2017)
Module Delivered in	5 programme(s)
Module Coordinator:	Sean McSweeney
Module Author:	Ignacio Castineiras
Module Description:	A NoSQL database provides a mechanism to store and retrieve data not modeled using tabular relations traditionally found in relational databases. In this module, the learner will survey the main data models used in NoSQL databases, with a special emphasis on models that support distributed data architectures. In addition, the learner will be equipped with the skills to store and query information in a NoSQL database.
Learning Outcomes	
<i>On successful completion of this module the learner will be able to:</i>	
LO1	Discuss how NoSQL data models and approaches can be applied to address challenges with Big Data.
LO2	Compare and contrast the ACID vs. BASE approaches (used respectively by relational and NoSQL data models) to store and retrieve data.
LO3	Survey the main key/value-based NoSQL families, identifying the uses cases most appealing to each of them.
LO4	Apply replication and sharding with a document-oriented NoSQL database to scale-out in a cluster.
LO5	Connect and synchronise a graph-based and a document-oriented NoSQL databases for a polyglot persistence-based solution.
Pre-requisite learning	
Module Recommendations <i>This is prior learning (or a practical skill) that is strongly recommended before enrolment in this module. You may enrol in this module if you have not acquired the recommended learning but you will have considerable difficulty in passing (i.e. achieving the learning outcomes of) the module. While the prior learning is expressed as named CIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).</i>	
Incompatible Modules <i>These are modules which have learning outcomes that are too similar to the learning outcomes of this module. You may not earn additional credit for the same learning and therefore you may not enrol in this module if you have successfully completed any modules in the incompatible list.</i>	
No incompatible modules listed	
Co-requisite Modules	
No Co-requisite modules listed	
Requirements <i>This is prior learning (or a practical skill) that is mandatory before enrolment in this module is allowed. You may not enrol on this module if you have not acquired the learning specified in this section.</i>	
No requirements listed	

Module Content & Assessment

Indicative Content

Introduction: Big Data.

New data storage and data processing context. Data Classification: 3 Vs. Scalability problem: Difficulties to capture, store, search, analyse and visualise the data. Emerging big data paradigm: New infrastructure (storage+compute cloud), data models and data manipulation techniques.

Traditional Relational Model.

Relational Databases Management Systems (RDBMS): De facto standard. Relational model: Transactional (ACID) properties, normalisation, expressive SQL language. RDBMS limitations for big data: Impedance mismatch, rigid schema, inability to deal with distributed huge data volumes.

NoSQL Alternative.

Paradigm arising to tackle problems RDBMS is not good at: Schema-less, high level data representation, scale-out distributed-based infrastructure. Wide range of data models: Pure key/value, column-based, document oriented and graph-based. Polyglot persistence. Lost of transactional properties: BASE approach.

Key/value DBs.

Pure key/value pairs data model: Simple and neat interface. Survey of advantages and disadvantages with respect to RDBMS. Use cases. Column-based DBs: More structured value pairs. Value gathering via supercolumns. Use cases.

Document-oriented DBs.

Most popular and flexible data model. JSON documents. Use cases. Consistency: Replica set. Scaling out: Sharding. Clusters, configuration nodes, shards, chunk of data, shard key range, balancing background operators. Complex querying: Aggregation framework, aggregation commands command pipelines.

Graph-based DBs.

Property graph data model: Nodes, relationships and properties. Use cases. Cypher query language. Connection and synchronisation with other data models.

Assessment Breakdown

	%
Course Work	100.00%

Course Work

Assessment Type	Assessment Description	Outcome addressed	% of total	Assessment Date
Project	An example project would be to store and query some information using a relational and two key/value-based NoSQL databases. As part of the assessment the student may be expected to determine which key/value family is more appealing based on the information being queried.	1,3	35.0	Week 7
Project	Given a distributed cluster to allocate a large dataset, store and query some information using a document-oriented and a graph-based NoSQL databases. Connect both databases for a polyglot persistence-based solution. Apply replication and sharding to demonstrate the scale-out approach followed. Assess the BASE transaction properties by disabling communication from/to some of the nodes.	2,4,5	35.0	Week 10
Short Answer Questions	This assessment will examine the theoretical content delivered in class.	1,2,3	30.0	Week 13

No End of Module Formal Examination

Reassessment Requirement

Coursework Only

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

The institute reserves the right to alter the nature and timings of assessment

Module Workload

Workload: Full Time				
<i>Workload Type</i>	<i>Workload Description</i>	<i>Hours</i>	<i>Frequency</i>	<i>Average Weekly Learner Workload</i>
Lecture	Lecture delivering theory underpinning learning outcomes.	2.0	Every Week	2.00
Lab	Practical computer-based lab supporting learning outcomes.	2.0	Every Week	2.00
Independent Learning	Independent Study.	3.0	Every Week	3.00
Total Hours				7.00
Total Weekly Learner Workload				7.00
Total Weekly Contact Hours				4.00

Workload: Part Time				
<i>Workload Type</i>	<i>Workload Description</i>	<i>Hours</i>	<i>Frequency</i>	<i>Average Weekly Learner Workload</i>
Lecture	Lecture delivering theory underpinning learning outcomes.	2.0	Every Week	2.00
Lab	Practical computer-based lab supporting learning outcomes.	2.0	Every Week	2.00
Independent Learning	Independent Study.	3.0	Every Week	3.00
Total Hours				7.00
Total Weekly Learner Workload				7.00
Total Weekly Contact Hours				4.00

Module Resources

Recommended Book Resources

- Pramod J. Sadalage and Martin Fowler 2013, *NoSQL Distilled: A Brief Guide to the Emerging World of Polyglot Persistence*, Addison-Wesley [ISBN: 9780321826626]
- John Sharp et. al. 2013, *Data Access for Highly-Scalable Solutions: Using SQL, NoSQL, and Polyglot Persistence*, Microsoft patterns & practices [ISBN: 9781621140306]

Supplementary Book Resources

- Kristina Chodorow 2013, *MongoDB: The Definitive Guide*, 2nd Ed., O'Reilly Media [ISBN: 9781449344689]
- Kyle Banker 2012, *MongoDB in Action*, Manning Publication [ISBN: 9781935182870]

This module does not have any article/paper resources

Other Resources

- Website: *MongoDB documentation*
<https://docs.mongodb.com/>
- Website: *Neo4j documentation*
<https://neo4j.com/docs/>
- Website: *Cassandra documentation*
<http://cassandra.apache.org/doc/latest/>

Module Delivered in

Programme Code	Programme	Semester	Delivery
CR_KSDEV_8	<u>Bachelor of Science (Honours) in Software Development</u>	4	Mandatory
CR_KDNET_8	<u>Bachelor of Science (Honours) in Computer Systems</u>	4	Mandatory
CR_KITMN_8	<u>Bachelor of Science (Honours) in IT Management</u>	8	Elective
CR_KCOMP_7	<u>Bachelor of Science in Software Development</u>	4	Mandatory
CR_KCOME_6	<u>Higher Certificate in Science in Software Development</u>	4	Mandatory