Module Details			
Module Code:	DATA C9001		
Full Title:	Data Architecture APPROVED		
Valid From::	Semester 1 - 2019/20 ( June 2019 )		
Language of Instruction:	English		
Duration:	2 Semesters		
Credits::	10		
Module Owner::	Peadar Grant		
Departments:	Unknown		
Module Description:	Students are familiarised with data and its storage within varied IT environments including cloud, onsite and legacy systems. A practical problem- based approach to relational, non-relational and allied data storage technologies is followed. Student analysts will interact with a wide variety of contemporary technologies and will specify suitable data storage systems for varied application domains.		

Module Learning Outcome		
On successful completion of this module the learner will be able to:		
#	Module Learning Outcome Description	
MLO1	Utilise industry-standard database systems for analytics workloads.	
MLO2	Design data storage components based on industry standard relational and non-relational databases.	
MLO3	Optimise storage and query performance for various database types	
MLO4	Construct appropriate interfacing for near-realtime heterogeneous data stores	
MLO5	Develop data architecture to store and process unstructured data in varied formats	
MLO6	Design suitable hardware and software solutions for data storage requirements in analytics-centric projects	
Per an and the language		

Pre-requisite learning

Module Recommendations 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 DkIT module(s) it also allows for learning (in another module or modules) which is equivalent to the learning specified in the named module(s).

No recommendations listed

#### Module Indicative Content

#### Data

Types of data: structured, semi-structured & unstructured data; files, streams and databases; four Vs of data; contemporary global data trends; modelling considerations; acquisition; storage and retrieval patterns; distributing; scaling; common file and stream data formats; compression.

#### IT environment

Analytics and transaction processing requirements; client/server data access patterns; analyst-client environment trends; shared file systems; server-centric database storage; mainframe data integration; storage devices; storage concepts (DAS/NAS/SAN); data centre, cloud and hybrid-cloud environments; object storage systems.

#### **Relational databases**

RDBMS system overview [PostgreSQL]; Application domains; tabular data (1-N-F); data types; data manipulation and querying using SQL; views; application query API; multi-table JOINS; foreign-key relationships; E-R modelling; geospatial data handling; user-defined functions; aggregate queries; transactions; ACID properties; replication; sharding; CAP theorem; RDBMS limitations.

#### Performance optimisation

#### Goals of optimisation; query planner and explanation; use of indices; materialised views; caching systems [Redis].

#### Non-relational databases

NoSQL characteristics; concept of BASE; implicit/explicit schema; problem-based practical application of range of non-relational database solutions to domain-specific data: document stores [MongoDB], key/value stores [Riak], column stores [Cassandra], graph databases [Neo4J], LDAP directories [Active Directory]; design considerations; ad-hoc and programmatic querying; non-relational facilities within RDBMS systems; RDBMS integration; clustering.

#### Unstructured data

Challenges of unstructured data; key application areas; large-file storage solutions; Role of Full-text searching; ETL of file-based data; rich-format data challenges [PDF, DOCX]; RDBMS-based full text search capabilities and limitations; full-text search engines; integration with RDBMS and Document store systems.

Module Assessment			
Assessment Breakdown	%		
Course Work	100.00%		
Module Special Regulation			

#### Assessments

Full Time On Campus				
Course Work				
Assessment Type	Class Test	% of Total Mark	15	
Marks Out Of	0	Pass Mark	0	
Timing	Week 10	Learning Outcome	1,2,3,4	
Duration in minutes	0			
Assessment Description Class test incorporating practical and electron	nic quiz components			
Assessment Type	Continuous Assessment	% of Total Mark	35	
Marks Out Of	0	Pass Mark	0	
Timing	End-of-Semester	Learning Outcome	1,2,3,4	
Duration in minutes	0			
Assessment Description Design and implementation of data storage s	ystem.			
Assessment Type	Class Test	% of Total Mark	15	
Marks Out Of	0	Pass Mark	0	
Timing	Week 10	Learning Outcome	1,2,5,6	
Duration in minutes	0			
Assessment Description Class test incorporating practical and electron	nic quiz components			
Assessment Type	Continuous Assessment	% of Total Mark	35	
Marks Out Of	0	Pass Mark	0	
Timing	End-of-Semester	Learning Outcome	1,2,5,6	
Duration in minutes	0			
Assessment Description Data Project 2 - A cross-module project end of Here, students will be encouraged to use reg	of semester project where students will design an ression and time series model for processing and	nd construct data storage system to efficiently ext d analysing data to make informed predictions.	ract the raw data and store the processed data.	
No Project				
No Practical				
No Final Examination				

### Part Time On Campus

Course Work				
Assessment Type	Class Test	% of Total Mark	15	
Marks Out Of	0	Pass Mark	0	
Timing	Week 10	Learning Outcome	1,2,3,4	
Duration in minutes	0			
Assessment Description Class test incorporating practical and electronic quiz components				
Assessment Type	Continuous Assessment	% of Total Mark	35	
Marks Out Of	0	Pass Mark	0	
Timing	End-of-Semester	Learning Outcome	1,2,3,4	
Duration in minutes	0			
Assessment Description Design and implementation of data storage system				
Assessment Type	Class Test	% of Total Mark	15	
Marks Out Of	0	Pass Mark	0	
Timing	Week 10	Learning Outcome	1,2,5,6	
Duration in minutes	0			
Assessment Description Class test incorporating practical and electronic quiz components				

#### Continuous Assessment 0 End-of-Semester

0

% of Total Mark	
Pass Mark	
Learning Outcome	

35 0 1,2,5,6

## Duration in minutes

Assessment Description Data Project 2 - A cross-module project end of semester project where students will design and construct data storage system to efficiently extract the raw data and store the processed data. Here, students will be encouraged to use regression and time series model for processing and analysing data to make informed predictions.

No Project

No Practical

No Final Examination

#### Reassessment Requirement

No repeat examination Reassessment of this module will be offered solely on the basis of coursework and a repeat examination will not be offered.

Reassessment Description Reassessment will consist of one design & implementation project covering and one class test covering both semesters' work.

Module Workload					
Workload: Full Time On Campus					
Workload Type	Contact Type	Workload Description	Frequency	Average Weekly Learner Workload	Hours
Practical	Contact	Practical lab session	Every Week	3.00	3
Independent Study	Non Contact	Practice with technologies studied in class	Every Week	4.00	4
Directed Reading	Non Contact	Lecturer-recommended supporting texts	Every Week	1.00	1
Total Weekly Learner Workload				8.00	
				Total Weekly Contact Hours	3.00
Workload: Part Time On	Campus				
Workload Type	Contact Type	Workload Description	Frequency	Average Weekly Learner Workload	Hours
Practical	Contact	Practical lab session	Every Week	3.00	3
Independent Study	Non Contact	Practice with technologies studied in class	Every Week	4.00	4
Directed Reading	Non Contact	Lecturer-recommended supporting texts	Every Week	1.00	1
	Total Weekly Learner Workload				8.00
				Total Weekly Contact Hours	3.00

# Module Resources Recommended Book Resources Connolly, Thomas & Begg, Carolyn. (2015), Database Systems, 6th. Addison Wesley. Pramod J. Sadalage and Martin Fowler. (2012), NoSQL Distilled, Addison Wesley. Luc Perkins, Eric Redmond, Jim Wilson. (2018), Seven Databases in Seven Weeks, 2nd. This module does not have any article/paper resources Other Resources Online manual, PostgreSQL 11 reference manual, https://www.postgresql.org/docs/11/index\_html Online manual, Riak database manual, https://docs.riak.com Online manual, Redis documentation, https://redis.io/documentation, https://redis.io/documentation, https://redis.io/documentation, https://redis.io/documentation, https://locssandra.apache.org/doc/latest/