

H7DA: Data Architecture

Module Code:	H7DA
Long Title	Data Architecture APPROVED
Title	Data Architecture
Module Level:	LEVEL 6
EQF Level:	5
EHEA Level:	Short Cycle
Credits:	5
Module Coordinator:	Paul Stynes
Module Author:	Paul Stynes
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	Master's degree in a computing or cognate discipline. May have industry experience also.
Learning Outcomes	
<i>On successful completion of this module the learner will be able to:</i>	
#	Learning Outcome Description
LO1	Compare different data stores, data models, query languages, data encoding techniques and knowledge discovery techniques
LO2	Summarise the constraints and trade-offs of a distributed shared-nothing architecture that is involved in the storage and retrieval of data
LO3	Analyse and design a data application architecture that integrates multiple disparate data systems that are optimised for different access patterns
LO4	Collaboratively implement an application data architecture
Dependencies	
Module Recommendations	
No recommendations listed	
Co-requisite Modules	
No Co-requisite modules listed	
Entry requirements	Learners should have attained the knowledge, skills and competence gained from stage 1 of the BSc (Hons) in Data Science

H7DA: Data Architecture

Module Content & Assessment			
Indicative Content			
Module Introduction Reliable, Scalable and Maintainable Applications Fundamentals of data systems. Reliability–H/D, S/W and Human errors.. Scalability-Load, Performance.. Maintainability- Operability, Simplicity and Evolvability.			
Data Models and Query Languages Relational Vs Document models.. Query Languages for Data.. Graph-Like Data Models.			
Storage and Retrieval Data Structures.. Transaction Processing or Analytics.Column-Oriented Storage.			
Encoding and Evolution Formats for Encoding Data.. Modes of Data Flow..			
Replication Leaders and Followers.Problems with Replication Lag.. Multi-Leader Replication. . Leaderless Replication.			
Partitioning and Transactions Partitioning and Replication.. Partitioning of Key-Value Data.. Partitioning and Secondary Indexes.. Rebalancing Partitions.. Request Routing.. Transaction.. Weak Isolation Levels.. Serializability.			
The trouble with Distributed Systems Faults and Partial Failures. Unreliable Networks. Unreliable Clocks.			
Consistency and Consensus Consistent Guarantees.Linearizability.Ordering Guarantees. Distributed Transactions and Consensus.			
Batch Processing Batch Processing with Unix Tools.MapReduce and Distributed Filesystems.			
Stream Processing Transmitting Event Streams.Database and Streams.Processing Systems.			
The future of Data Systems Data Integration. Unbundling Databases. Predictive Analytics			
Knowledge discovery KDD. CRISP-DM			
Assessment Breakdown			%
Coursework			60.00%
End of Module Assessment			40.00%
Assessments			
Full Time			
Coursework			
Assessment Type:	Continuous Assessment	% of total:	Non-Marked
Assessment Date:	n/a	Outcome addressed:	1,2,3,4
Non-Marked:	Yes		
Assessment Description: Ongoing independent and group class activities and feedback.			
Assessment Type:	Project	% of total:	60
Assessment Date:	n/a	Outcome addressed:	1,2,3,4
Non-Marked:	No		
Assessment Description: Students will work collaboratively to implement an application data architecture that integrates a variety of data systems with different data models and different query languages to access the data. Learners will integrate the data and apply data encoding techniques and knowledge discovery techniques to gain insight from the data.			
Assessment Type:	Easter Examination	% of total:	40
Assessment Date:	n/a	Outcome addressed:	1,2,3
Non-Marked:	No		
Assessment Description: The examination may include a mixture of theoretical, applied and interpretation questions.			
No End of Module Assessment			
No Workplace Assessment			
Reassessment Requirement			
Repeat examination <i>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.</i>			
Reassessment Description The repeat strategy for this module is a terminal assessment. Students will be afforded an opportunity to repeat the assessment at specified times throughout the year and all learning outcomes will be assessed in the repeat assessment.			

H7DA: Data Architecture

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	Per Semester	2.00
Tutorial	Other hours (Practical/Tutorial)	12	Per Semester	1.00
Independent Learning	Independent learning (hours)	89	Per Semester	7.42
Total Weekly Contact Hours				3.00

Module Resources	
Recommended Book Resources	
Kleppmann, M.. (2017), Designing Data-Intensive Applications – The big ideas behind reliable, scalable and maintainable systems, California, O'Reilly Media Inc.	
Supplementary Book Resources	
Adkins, S., Belamaric, J., Giersch, V., Makogon, D. & Robinson, J.. (2015), OpenStack Cloud Application Development, Wrox.	
This module does not have any article/paper resources	
Other Resources	
[Website], http://hadoop.apache.org/ [Website], https://kafka.apache.org/intro [Website], https://thrift.apache.org/ [Website], http://storm.apache.org/ [Website], https://parquet.apache.org/ [Website], https://www.elastic.co/products/elasticsearch [Website], https://memcached.org/ [Website], https://www.mongodb.com/ [Website], https://neo4j.com/product/ [Website], https://developers.google.com/protocol-buffers/ [Website], https://www.w3.org/RDF/ [Website], https://redis.io/ [Website], http://lucene.apache.org/solr/ [Website], http://www.w3.org/TR/rdf-sparql-query/	
Discussion Note:	