

H8DFA: Databases for Analytics

Module Code:	H8DFA
Long Title	Databases for Analytics APPROVED
Title	Databases for Analytics
Module Level:	LEVEL 8
EQF Level:	6
EHEA Level:	First Cycle
Credits:	10
Module Coordinator:	EUGENE O'LOUGHLIN
Module Author:	ORLA LAHART
Departments:	School of Computing
Specifications of the qualifications and experience required of staff	PhD or MSc degree required in computer science or cognate discipline. Experience of lecturing in the field of Computing/ Databases. 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	Analyse and understand the importance of data models for relational and non-relational databases.
LO2	Develop conceptual and logical models for relational databases using knowledge of entity-relationship diagrams and normalization.
LO3	Codify and retrieve data using structured query language to manipulate databases through effective reporting.
LO4	Investigate the use of the data warehousing for analytics using dimensional models.
LO5	Explore and apply NoSQL databases for the Big data storage as a solution.
LO6	Construct data lakes for storage, processing and data analytics.
Dependencies	
Module Recommendations	
No recommendations listed	
Co-requisite Modules	
No Co-requisite modules listed	
Entry requirements	Internal to the programme

H8DFA: Databases for Analytics

Module Content & Assessment			
Indicative Content			
Database Concepts and DBMS Data, Databases, File based systems, DBMS environment, Data Redundancy, Data Anomalies, Database System Environment, Functions of DBMS, Database Professional Career and Concept of views.			
Data and Database Modelling Data modelling, Relational Models, Object-relational Model and Object Models, Importance of Relational databases and comparison of database models.			
Entity Relationship (ER) Modelling Concepts and terminology in Entity–Relationship (ER) modelling, Diagrammatic technique using Unified Modelling Language (UML), Challenges associated with ER models, Build ER model based on requirements specifications.			
Logical Modelling (Normalization) The purpose of normalization, Designing a relational database, Potential problems associated with redundant data, Concept and Characteristics of functional dependency, Process of normalization, First Normal Form (1NF), Second Normal Form (2NF), and Third Normal Form (3NF).			
Physical Modelling using Data Definition Language (DDL) Data types supported by SQL, Define integrity constraints, Integrity enhancement feature in the CREATE and ALTER TABLE statements, Create and delete views using SQL, Operations on views, Advantages and disadvantages of views, GRANT and REVOKE statements as a level of security, Transactions (COMMIT and ROLLBACK) in SQL.			
Structured Query Language (SQL) for Data retrieval Importance of SQL, Writing of SQL commands, Retrieve data using the SELECT statement, Build SQL statements, use the WHERE clause to retrieve rows that satisfy various conditions, sort query results using ORDER BY, use the aggregate functions of SQL and group data using GROUP BY.			
Structured Query Language (SQL) for Data retrieval Writing of SQL commands for complex queries, Retrieve data using the SELECT statement from multiple tables, Build SQL statements (Continuation from previous week), use subqueries, join tables together, perform set operations (UNION, INTERSECT, EXCEPT) and Perform database updates using INSERT, UPDATE, and DELETE.			
Indexing and Hashing Introduction to Indexing, Types of indexing, Ordered Indices, Static Hashing, Dynamic Hashing, Comparison of Ordered Indexing and Hashing, Index Definition in SQL and Tree Structures.			
Data Warehousing Concepts Concepts and benefits of data warehousing, Online transaction processing (OLTP), Architecture and main components, Tools and technologies, Concept of a data mart and the main reasons for implementing a data mart, Benefits of Data warehousing.			
Data Warehousing Design and Data Partitioning Designing Data Warehouses, Two main methodologies: Inmon's Corporate Information Factory (CIF) and Kimball's Business Dimensional Lifecycle, Principles and stages associated with Kimball's Business Dimensional Lifecycle, Dimensional Modelling stage of Kimball's Business Dimensional Lifecycle, Issues associated with the development of a data warehouse.			
Non-relational Databases and Analytics NoSQL databases, Capabilities of NoSQL Technologies, E-commerce/ Social media Applications, Motivations for NoSQL Databases, The CAP Theorem, Consequences of The CAP Theorem, Data Management with Distributed Databases, ACID and BASE Characteristics, NoSQL Classifications, Benefits of NoSQL Databases.			
Data Lakes for SQL and NoSQL databases Introduction to Data Lakes, Data Lakes for enterprises and works, Differentiate between Data Lake and Data Warehouse, Lambda Architecture for Data Lakes, Batch layer, Speed layer, Data Storage layer, Serving layer, Data Acquisition of Batch, Data Processing and Data Store using Apache Hadoop			
Assessment Breakdown			%
Coursework			50.00%
End of Module Assessment			50.00%
Assessments			
Full Time			
Coursework			
Assessment Type:	Test	% of total:	25
Assessment Date:	n/a	Outcome addressed:	1,2
Non-Marked:	No		
Assessment Description: Case Study for ER Modelling or Normalization, Online Quiz for Conceptual understanding of database concepts.			
Assessment Type:	Test	% of total:	25
Assessment Date:	n/a	Outcome addressed:	3
Non-Marked:	No		
Assessment Description: A scenario for the relational database will be provided along with details for the relations and attributes. The students will transform this case study information into practical database, insert data into the database tables, perform operations for data manipulation language and retrieve important information through SQL queries.			
End of Module Assessment			
Assessment Type:	Terminal Exam	% of total:	50
Assessment Date:	End-of-Semester	Outcome addressed:	1,4,5,6
Non-Marked:	No		
Assessment Description: Terminal examination will be based on understanding of Database Concepts and relational models, Dimensional Modelling of Data warehousing, NoSQL databases and application of data lakes in business requirements.			
No Workplace Assessment			
Part Time			
No Coursework			

End of Module Assessment			
Assessment Type:	Terminal Exam	% of total:	0
Assessment Date:	End-of-Semester	Outcome addressed:	1,4,5,6
Non-Marked:	No		
Assessment Description: Terminal examination will be based on Conceptual understanding of Database Concepts, Dimensional Modelling of Data warehousing and usage of data warehousing for analytics, NoSQL databases for Bid data and the role of data lakes for the analytics.			
Assessment Type:	Terminal Exam	% of total:	0
Assessment Date:	End-of-Semester	Outcome addressed:	1,4,5,6
Non-Marked:	No		
Assessment Description: Terminal examination will be based on understanding of Database Concepts and relational models, Dimensional Modelling of Data warehousing, NoSQL databases and application of data lakes in business requirements.			
Assessment Type:	Terminal Exam	% of total:	0
Assessment Date:	End-of-Semester	Outcome addressed:	1,4,5,6
Non-Marked:	No		
Assessment Description: Terminal examination will be based on understanding of Database Concepts and relational models, Dimensional Modelling of Data warehousing, NoSQL databases and application of data lakes in business requirements.			
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 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.			

H8DFA: Databases for Analytics

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	
<i>Recommended Book Resources</i>	
<p>Hoffer, Venkataraman & Topi. (2019), Modern Database Management, 13th edition. Pearson, p.592, [ISBN: 9780134792279].</p> <p>Thomas Connolly, Thomas M. Connolly, Carolyn E. Beg. (2015), Database Systems, 6th edition. Addison-Wesley, p.1440, [ISBN: 9780132943260].</p>	
<i>Supplementary Book Resources</i>	
<p>Carlos Coronel, Steven Morris. (2018), Database Systems: Design, Implementation, & Management, 13th Edition. Cengage Learning, p.816, [ISBN: 978-1-337-62790-0].</p> <p>Dan Sullivan. (2015), NoSQL for Mere Mortals, Pearson Education, p.510, [ISBN: 9780134023212].</p> <p>Alex Gorelik. (2019), The Enterprise Big Data Lake, O'Reilly, p.200, [ISBN: 9781491931554].</p> <p>John Viescas, Douglas J Steele, Ben Clothier. (2017), Effective SQL, 1st edition. Addison-Wesley Professional, p.300, [ISBN: 9780134578897].</p>	
<i>Recommended Article/Paper Resources</i>	
<p>Codd E.F. (1982), The 1981 ACM Turing Award Lecture: Relational database: A practical foundation for productivity, Comm. ACM, 25, https://dl.acm.org/doi/10.1145/1283920.1283937</p>	
<i>Supplementary Article/Paper Resources</i>	
<p>Codd E.F.. (1970), A Relational Model of Data for Large Shared Data Banks, IBM Research Laboratory, San Jose, California, https://www.seas.upenn.edu/~zives/03f/cis550/codd.pdf</p> <p>Embley, D. W. (1989), NFQL: The Natural Forms Query Language, ACM Transactions on Database Systems, 1989, https://dl.acm.org/doi/10.1145/63500.64125</p>	
<i>Other Resources</i>	
<p>[Website], MySQL Tutorial, https://www.mysqltutorial.org</p> <p>[Website], DataCamp https://www.datacamp.com/courses/introduction-to-sql.</p> <p>[Website], MySQL Tutorial https://www.tutorialspoint.com/mysql/index.htm.</p> <p>[Website], Draw.io https://app.diagrams.net/.</p> <p>[Website], MongoDB https://www.mongodb.com/nosql-explained.</p> <p>[Website], NoSQL https://hostingdata.co.uk/nosql-database/.</p>	
Discussion Note:	