

Gültig für 2022.HS

| <b>Modulbezeichnung: Data Engineering</b>   |   |
|---|---|
| Modulkürzel   | w.BA.XX.3DE-WIN.XX  |
| Beschreibung des Moduls   | Data Engineering als Methodik und Aufgabenbereich umfasst die ersten Schritte des Datenlebenszyklus von der Generierung und Beschaffung von Daten, über die Speicherung, die Übertragung sowie Transformation bis zur automatisierten Bereinigung mit Monitoringsystemen und Real-Time Analytics. In diesem Modul lernen die Studierenden die einzelnen Phasen dieses Lebenszyklus kennen und welche technischen und geschäftlichen Überlegungen bei der Wahl der Systeme eine Rolle spielen. Weiter wenden die Studierenden selbst gängige Methoden und Techniken in den einzelnen Phasen des Lebenszyklus an.   |
| Studiengang und Vertiefungsrichtung   | § Wirtschaftsinformatik<br>§ Wirtschaftsinformatik - Vertiefung in Business Information Systems<br>§ Wirtschaftsinformatik - Vertiefung in Data Science   |
| Rechtliche Grundlagen   | Studienordnung BSc vom 29.01.2009, Anhang zur Studienordnung für die Bachelorstudiengänge Betriebsökonomie, Wirtschaftsinformatik und Wirtschaftsrecht und Angewandtes Recht, erstmals beschlossen am 12.05.2009  |
| Modulkategorie  | <b>Modultyp:</b><br>Pflichtmodul <b>Studienabschnitt:</b><br>Hauptstudium   |
| ECTS  | 6   |
| Verantwortliche OE  | W Institut für Wirtschaftsinformatik  |
| Modulverantwortung  | Maria Pelli (pell)  |
| Modulverantwortung Stellvertretung  | Christian Hitz (hitz)   |
| Spezifische Vorkenntnisse   | - Scientific Programming (w.BA.XX.3SP-WIN.XX)<br>- Data Management (w.BA.XX.3DM-WIN.XX)   |
| Beitrag des Moduls für Learning Goals des Studiengangs (durch das Modul betroffene) | § Fachkompetenz<br>§ Methodenkompetenz<br>§ Sozialkompetenz<br>§ Selbstkompetenz  |
| Beitrag des Moduls für Learning Objectives des Studiengangs                         | Fachkompetenz<br>§ Theorie- & Praxisrelevante Fachinhalte wissen & verstehen<br>§ Theorie- & Praxisrelevante Fachinhalte anwenden, analysieren und verknüpfen<br>§ Theorie- & Praxisrelevante Fachinhalte evaluieren<br>Methodenkompetenz<br>§ Problemlösung & Kritisches Denken<br>§ Wissenschaftliche Methoden<br>§ Arbeitsmethoden, -techniken & -verfahren<br>§ Nutzung von Informationen<br>§ Kreativität & Innovation<br>Sozialkompetenz<br>§ Schriftliche Kommunikation<br>§ Mündliche Kommunikation<br>§ Kooperation im Team & Umgang mit Konflikten<br>§ Interkulturalität & Perspektivenübernahme<br>Selbstkompetenz<br>§ Selbstmanagement & Selbstreflexion<br>§ Ethische & Soziale Verantwortung<br>§ Lernen & Veränderung  |
| Lernziele des Moduls  | Die Studierenden...<br>§ Verstehen und Abgrenzen des Data Engineering als Methodik und Aufgabenbereich im Kontext verwandter Disziplinen und Methodiken<br>§ Kennen, Verstehen und Analysieren der wichtigsten Datenproblematiken und Herausforderungen im Bereich des Data Engineering<br>§ Kennen und Verstehen des Lebenszyklus des Data Engineering und dessen Unterströmungen<br>§ Kennen der wichtigsten Architekturen im Lebenszyklus des Data Engineering<br>§ Kennen der wichtigsten technischen Überlegungen bei der Wahl und Evaluation von Quellsystemen und deren Eigenschaften<br>§ Kennen und Anwenden der gängigsten Methoden im Bereich des Data Wrangling<br>§ Kennen der Unterschiede sowie Vor- und Nachteile verschiedener Architekturen innerhalb des Lebenszyklus des Data Engineering<br>§ Kennen, Verstehen und Evaluieren des Kompromisses zwischen Konsistenz, Verfügbarkeit und Ausfalltoleranz |

|  |   |  |                                |
|--|---|--|--------------------------------|
|  | § Kennen und Verstehen der gängigsten Speicherformen im Bereich des Distributed Data Storage<br>§ Kennen wichtiger Data Ingestion - Methoden und die dafür erforderlichen Dateneigenschaften<br>§ Kennen und Anwenden von Ansätzen des Parallel und Cluster Computings<br>§ Kennen und Gegenüberstellen der wichtigsten Überlegungen der Rentabilität und des Geschäftswertes bei der Wahl von Transformationsmethoden und deren Einordnung im Lebenszyklus des Data Engineering<br>§ Kennen und Anwenden der wichtigsten Methoden im Bereich des Parsings und der Datenzusammenführung<br>§ Kennen und Anwenden der gängigsten Methoden zur Erkennung und Behandlung von Anomalien (Data Cleansing)<br>§ Kennen der wichtigsten analytischen Systeme und deren Zusammenhänge mit Data Engineering<br>§ Kennen und Anwenden der Methodik des Reverse ETL<br>§ Kennen und Anwenden von Systemen für Monitoring und Real-time Analytics für automatisierte Datenaufbereitung (Data Scrubbing) |  |                                |
| Inhalte des Moduls                     | § Einführung, historischer Überblick, Definition und Abgrenzung von Data Engineering<br>§ Data Engineering Lifecycle: Übersicht des Lebenszyklus und dessen Unterströmungen<br>§ Architekturen im Data Engineering Lifecycle: Übersicht über Referenzarchitekturen, Schema Evolution<br>§ Generierung und Beschaffung von Daten: Evaluation von Quellsystemen und technische Voraussetzungen<br>§ Storage: Übersicht über Speicherformen, Kompromiss der Konsistenz, Verfügbarkeit und Partitionstoleranz, Distributed Data Storage<br>§ Ingestion: Datenfrequenzen, Konzepte und Architekturen<br>§ Transformation: ROI in Bezug auf Transformation und dessen Einordnung im Data Engineering Lifecycle, Parsing, Formen der Zusammenführung von Daten und deren Voraussetzungen, Erkennung und Behandlung von Anomalien<br>§ Serving: Übersicht über analytische Systeme und deren Zusammenhänge mit Data Engineering, Reverse ETL, Monitoring und Real-Time Analytics für Data Scrubbing |  |                                |
| Verknüpfung zu anderen Modulen         | -   |  |                                |
| Unterrichtsmethoden                    | § Lehrvortrag<br>§ Fallstudien<br>§ Übungen<br>§ Literaturstudium   | <b>Eingesetzte Sozialformen:</b><br>§ Einzelarbeit<br>§ Partnerarbeit<br>§ Gruppenarbeit |                                |
| Digitale Lernressourcen                | § Reader<br>§ Lehrvideos<br>§ Übungs- und Anwendungsaufgaben (inkl. Lösungen)<br>§ Fallstudien (inkl. Lösungen)   |  |                                |
| Unterrichtsgliederung                  | <b>Kontaktstudium</b>   | <b>Begleitetes Selbststudium</b>   | <b>Autonomes Selbststudium</b> |
|  | Grossklasse   | 28 h   | -                              |
|  | Kleinklasse   | 28 h   | 39 h                           |
|  | Gruppenunterricht   | -  | -                              |
|  | Praktikum   | -  | -                              |
|  | Seminar   | -  | -                              |
|  | <b>Total</b>  | <b>56 h</b>  | <b>39 h</b>                    |
|  |   |  | <b>85 h</b>                    |
| Leistungsnachweise                     |   |  |                                |
|  | <b>Modulendprüfung</b>  | <b>Form</b>  | <b>Dauer (Min.)</b>            |
|  | Schriftliche Prüfung  | open book  | 90                             |
|  | <b>Hilfsmittel</b>  | spezifizierter Taschenrechner gem. "Merkblatt Hilfsmittel"                               | mit Diktionär                  |
|  | <b>Gewichtung</b>   |  | 100,00 %                       |
|  | <b>Andere</b>   | <b>Bewertung</b>   | <b>Dauer (Min.)</b>            |
|  | -   | -  | -                              |
| Präsenzverpflichtung im Kontaktstudium | Zwingende Präsenzzeit: Keine  |  |                                |
| Unterrichts- und Prüfungssprache       | Deutsch   |  |                                |
| Pflichtliteratur                       | Reis, J. & Housley, M. (2022). Fundamentals of Data Engineering. Sebastopol: O'Reilly Media, Inc. ISBN 978-1098108304.  |  |                                |
| Ergänzende Literatur                   | -   |  |                                |
| Bemerkungen                            | Es wird ein Notebook vorausgesetzt. Die prüfungsrelevante und optionale Literatur kann in Deutsch oder Englisch verfasst sein.  |  |                                |