

Modulname	<b>Scalable Data Engineering</b>
Modulnummer	INF-25a-Ma-FSA-SDE
Verantwortliche Dozentin bzw. verantwortlicher Dozent	Prof. Dr.-Ing. Wolfgang Lehner wolfgang.lehner@tu-dresden.de
Qualifikationsziele	Die Studierenden kennen erweiterte Konzepte der Analyse großer Datenbestände im Bereich der Data Science. Zudem kennen sie den gesamten Prozess von der Erfassung bis hin zur Auswertung von Datenbeständen unter Verwendung geeigneter Konzepte und Werkzeuge. Sie verfügen über Wissen zu den einzelnen Schritten für die Erfassung von Daten (data extraction), der Inspektion (data profiling), der Ermittlung von Datenqualität (data scrubbing), der für analytische Umgebungen notwendigen Modellierung (data warehousing, data lakes) und schließlich der Klassen von Analyseverfahren zum Erreichen der Qualifikationsziele. Die Studierenden sind in der Lage, die Komplexität der Datenanalyse einzuschätzen und entsprechende methodische als auch technische Lösungen auszuwählen und deren Vor- und Nachteile zu bewerten. Sie kennen den praktischen Umgang mit einzelnen Werkzeugen für die unterschiedlichen Prozessschritte.
Inhalte	Inhalt des Moduls ist der gesamte Prozess der Datenanalyse von der Erfassung umfangreicher Datenbestände bis hin zur Diskussion einzelner Analyseverfahren, wie beispielsweise das Data Profiling, das Schema Discovery, die Bewertung der Datenqualität sowie Modellierungsalternativen für analytische Zugriffsmuster. Weitere Inhalte sind systematische Analysealgorithmen aus unterschiedlichen Verfahrensklassen hinsichtlich der Befriedigung des Informationsbedürfnisses des Benutzers und der systemtechnischen Auswirkungen.
Lehr- und Lernformen	Das Modul umfasst Vorlesungen im Umfang von 2 SWS, Übungen im Umfang von 2 SWS und das Selbststudium. Die Lehrsprache der Vorlesungen und der Übungen kann Deutsch oder Englisch sein und wird jeweils zu Semesterbeginn von der Dozentin bzw. dem Dozenten konkret festgelegt und in der jeweils üblichen Weise bekannt gegeben.
Voraussetzungen für die Teilnahme	Im Diplomstudiengang Informatik werden die im Modul INF-25-Ba-DMF Data Management Foundations zu erwerbenden Kompetenzen vorausgesetzt.
Verwendbarkeit	Das Modul ist im Diplomstudiengang Informatik im Hauptstudium ein Wahlpflichtmodul im Fachgebiet Systems Architecture, das nach Maßgabe der Anlage 2 zur Prüfungsordnung zu wählen ist. Das Modul ist im Masterstudiengang Computer Science jeweils ein Wahlpflichtmodul im Open Track im Fachgebiet Systems Architecture sowie der Ergänzung, das nach Maßgabe der Anlage 2 zur Prüfungsordnung zu wählen ist. Das Modul kann im Masterstudiengang Computer Science nur einmal gewählt werden. Das Modul kann im Masterstudiengang Computer Science nicht

	<p>gewählt werden, wenn dieses oder ein wesentlich inhaltsgleiches Modul aus einem Studiengang mit dem die Zugangsvoraussetzungen nach § 3 der Studienordnung erfüllt wurden, bereits absolviert wurde. Das Modul schafft die Voraussetzungen für die Module, die es unter Voraussetzungen für die Teilnahme benennen.</p>
Voraussetzungen für die Vergabe von Leistungspunkten	<p>Die Leistungspunkte werden erworben, wenn die Modulprüfung bestanden ist. Die Modulprüfung besteht aus einer Klausurarbeit von 90 Minuten Dauer. Die Prüfungssprache kann Deutsch oder Englisch sein und wird jeweils zu Semesterbeginn von der Dozentin bzw. dem Dozenten konkret festgelegt und in der jeweils üblichen Weise bekannt gegeben.</p>
Leistungspunkte und Noten	<p>Durch das Modul können 6 Leistungspunkte erworben werden. Die Modulnote entspricht der Note der Prüfungsleistung.</p>
Häufigkeit des Moduls	<p>Das Modul wird jedes Wintersemester angeboten.</p>
Arbeitsaufwand	<p>Der Arbeitsaufwand beträgt insgesamt 180 Stunden.</p>
Dauer des Moduls	<p>Das Modul umfasst ein Semester.</p>
Begleitliteratur	<p><a href="http://www.sde.db-tu-dresden.de">http://www.sde.db-tu-dresden.de</a></p>