

Lehrende/r	Prof. Dr. Uta Störl	Modulverantwortliche/r	Prof. Dr. Uta Störl
	Dauer des Moduls ein Semester	ECTS 10	Workload 300 Stunden
			Häufigkeit in jedem Semester
Lehrveranstaltung(en)	Data Engineering für Data Science		
Detaillierter Zeitaufwand	Bearbeiten der Lektionen: 160 Stunden Bearbeitung der Einsendeaufgaben inkl. Verarbeitung des Korrektur-Feedbacks: 80 Stunden Wiederholung und Prüfungsvorbereitung, Prüfung: 60 Stunden		
Qualifikationsziele	Studierende erwerben in diesem Modul Kenntnisse über Data Engineering für Data Science. Dies umfasst die vertiefenden Konzepte einer Programmiersprache (Python) und einer Datenbanksprache (SQL) für die Datenaufbereitung und -validierung sowie die Konzepte von Big-Data-Infrastrukturen. Studierende erlangen Kompetenzen und Verantwortungsbewusstsein für die Arbeit mit großen Datenmengen (Big Data).		
Inhalte	<ul style="list-style-type: none"> <li>- Methoden und Algorithmen im Kontext der Verarbeitung von großen Datenmengen (Big Data)</li> <li>- Voraussetzungen und Herausforderungen von Data Wrangling und Data Quality</li> <li>- Data Wrangling und Datenanalysen mit Python und SQL</li> <li>- Verteilte und parallele Big-Data-Infrastrukturen (Hadoop, Spark etc.)</li> <li>- Big-Data-Referenzarchitekturen</li> <li>- Verteilte nicht-relationale Datenbanksysteme (NoSQL-Datenbanksysteme)</li> </ul>		
Inhaltliche Voraussetzung	Grundkenntnisse Datenbanksysteme und Programmierung,  Hardware (mindestens): <ul style="list-style-type: none"> <li>- 16 GB RAM</li> <li>- CPU mit 8 Cores</li> <li>- 4 GB freier Festplattenspeicher</li> </ul>		
Lehr- und Betreuungsformen	Lehrveranstaltungsmaterial Betreuung und Beratung durch Lehrende Einsendeaufgaben mit Korrektur und/oder Musterlösung internetgestütztes Diskussionsforum Lehrvideos Video-Meetings		
Anmerkung	Der Basistext muss vor Semesterbeginn beschafft werden. Basistext: Susan E. McGregor: Practical Python Data Wrangling and Data Quality, O'Reilly Media, Inc., 2021 ISBN: 9781492091509		
Formale Voraussetzung	keine		
Verwendung des Moduls	M.Sc. Data Science M.Sc. Informatik M.Sc. Praktische Informatik		

## Prüfungsformen

Prüfung

Stellenwert  
der Note 1/12

Art der Prüfungsleistung

benotete zweistündige  
Prüfungsklausur

Voraussetzung

Von den Einsendeaufgaben zu den Lektionen  
2-5 müssen mindestens die Einsendeaufgaben  
zu einer Lektion bestanden sein.