

Data Engineering

Modulbezeichnung	Data Engineering
Modulverantwortlicher	J. Stöß
Modulniveau	Bachelor
EDV-B.	DSCB330
Modulumfang (ECTS)	6
Semester	3
Lernziele & Kompetenzen	<p>Das Modul soll den Studenten grundlegende Kenntnisse des Data Engineering als ingenieurwissenschaftliche Herangehensweise bei der Generierung, Speicherung, Pflege, Aufbereitung, Anreicherung und Weitergabe von Daten vermitteln. Die Studenten sollen fundamentale Konzepte, Methoden und Realisierungen von Datenmanagement und Datenanalyse im Kontext der Data Science kennen und anwenden lernen.</p> <p>Außerdem lernen die Studenten Techniken und Methoden zur Realisierung datenintensiver und datengetriebener Anwendungssysteme für die Data Science kennen. Sie setzen sich mit Architekturen und Komponenten daten- und wissensgetriebener Anwendungen und können an der Realisierung solcher Systeme mitwirken.</p> <p>Fachliche Kompetenzen/Lernergebnisse</p> <ul style="list-style-type: none">• Die Studenten kennen und verstehen die Zielsetzung und Prinzipien des Data Engineering in allen Bereichen der Data Science: bei der Vorverarbeitung, Aufbereitung, Analyse, Visualisierung von Daten, sowie konkret im Kontext tatsächlicher Anwendungen wie Berichtssysteme, Business Intelligence & Data Warehouses, Data Mining, oder Machine Learning.• Die Studenten kennen und verstehen die grundlegenden Herausforderungen und Aspekte des Data Engineering, und entwickeln ein Verständnis des gesamten "Data Life Cycle" im Kontext datenintensiver und datenwissenschaftlicher Softwaresysteme:<ul style="list-style-type: none">◦ <i>Datenmodellierung</i>: von konventionellen relationalen Modellen zu Ontologien, Graph-Modellen, RDF und anderen Modellierungssprachen oder Fragestellungen der Wissensrepräsentation◦ <i>Datensammlung und Datenstrukturierung</i>: Aspekte der Datensammlung, Datensäuberung, Datenintegrität, Datenherkunft, Datenfehler, Datenhygiene, usw.◦ <i>Datenintegration</i>: Konnektivität von Daten, Data Ingestion (ETL), und Dateninterfaces (IDL, Data Formats), Machine Learning Deployment und Solution Engineering◦ <i>Analytische Datenabfrage, Datenverarbeitung und Datensemantik</i>: Daten- und Analyseabfragen (SQL, SPARQL, MapReduce), Kategorien von Abfragesprachen (funktional, deklarativ, imperativ), Analytische Datenverarbeitung (Spark, MapReduce, Data Mining), Data Science Architekturen (Feature Engineering, Predictive Systems, Modeling und Machine Learning) Daten- und Transaktionssemantik (ACID, verteilte Transaktionen, NoSQL, CAP-Theorem)◦ <i>Datenhaltung, Datenspeicherung, Datenintegration</i>: Datenbanken und Dateisysteme, Service- und Cloud Computing, Service- und Cloud Computing, Streaming und Event-basierte Architekturen, Distributed File Systems & Spaltenorientierte Datenhaltung und Datenbanken, Search & Discovery◦ <i>Datenkodierung und Datenevolution</i>: Datentypen, Datenformate, Schemata, Schemakonflikte & auflösung, Kodierungen, Serialisierung, usw.◦ <i>Visualisierung: Dashboards, Real-Time Intelligence, ...</i>◦ <i>Big Data, Massendatenverarbeitung</i>: von relationalen zu NoSQL Architekturen, Volume/Velocity/Variety, Grenzen relationaler Datenverarbeitung, Skalierbare Datenverarbeitung mit Hadoop oder Spark, Data Lake Architekturen, ...◦ <i>Technische Aspekte datenintensiver Software- und Rechensysteme</i>:

Zuverlässigkeit, Skalierbarkeit, Datenverteilung, verteilte Datenverarbeitung, Wartbarkeit, Fehlertoleranz

- Die Studenten lernen tatsächliche Anwendungsfälle am praktischen Beispiel kennen und beschäftigen sich mit deren Realisierung und Implementierung:
 - Datenformate und Kodierung: JSON, XML, Protobuf, ...
 - Abfragesprachen: SQL, Python, SPARQL, ...
 - Datenbankbasierte Softwaresysteme & Data Warehouses
 - Data Mining & Knowledge Discovery
 - Wissenschaftliches Rechnen
- Die Studenten lernen ausgewählte Data Engineering und Data Science Referenzarchitekturen und Komponenten kennen und studieren diese aus dateningenieurwissenschaftlicher Perspektive
 - Big Data & NoSQL Systems
 - Graph Based Systems
 - Predictive Analytics / Machine Learning Systems
 - Visualization and Observability Platforms

Methodische Kompetenzen/Lernergebnisse

- Die Studenten kennen Konzepte und Methoden des Data Engineering und deren Rolle in den Datenwissenschaften,
- Die Studenten können diese Konzepte und Methoden auf konkrete Fragestellungen im Bereich Data Engineering anwenden, andere und alternative Herangehensweisen bewerten und vergleichen, sowie Handlungsempfehlungen fundiert ableiten.
- Grundlegende Methoden der Qualitätssicherung kennen und anwenden
- Die Studenten kennen Konzepte und Methoden zur Lösung von Data Engineering Problemen im Kontext der Datenwissenschaften
- Die Studenten können diese Konzepte und Methoden auf konkrete Fragestellungen im gesamten "Data Life Cycle" datenwissenschaftlicher Systeme anwenden, andere und alternative Herangehensweisen bewerten und vergleichen, sowie Handlungsempfehlungen fundiert ableiten.
- Grundlegende Methoden der Qualitätssicherung kennen und anwenden

Sozial- und Selbstkompetenz

- Zusammenarbeit in Teams zur Lösung kleinerer Probleme im Bereich des Programmierens
- Arbeitsplanung und -kontrolle
- Selbsteinschätzung

Lehr- und Lernform	Vorlesung mit integrierten Übungen
Assoziierte Module	-
Verwertbarkeit des Moduls	(s. Studiengangskonzept)
Inhaltliche Voraussetzungen	Datenbanken und Datenkunde 2
Voraussetzungen nach SPO	-
Prüfungsleistung	Klausur 90 Min.