

Modulbezeichnung: Big Data Project

Modulkürzel	t.BA.DS.PM4.20HS
ECTS Credits	4
Unterrichts- und Prüfungssprache	Deutsch
Verantwortliche OE	InIT
Modulverantwortung	Jonathan Fürst
Rechtliche Grundlagen	Die Modulbeschreibung ist neben Rahmenprüfungsordnung und Studienordnung Teil der Rechtsgrundlage. Sie ist verbindlich. Eine in der ersten Unterrichtswoche des Semesters schriftlich festgehaltene und kommunizierte Modulvereinbarung kann die Modulbeschreibung präzisieren. Die Modulvereinbarung ersetzt nicht die Modulbeschreibung.
Modulusprägung	Typ 4* 4 Lektionen Praktikum pro Semesterwoche und Halbkasse
Beschreibung des Moduls	Der Schwerpunkt dieses Moduls liegt darauf, praktische Erfahrungen im Umgang mit Big-Data-Problemen zu sammeln. Das Modul basiert auf den theoretischen Grundlagen von "Data Engineering 1" und "Data Engineering 2". Die Studierenden werden ausgewählte Themen dieser Grundlagenkurse analysieren und dann skalierbare Anwendungen unter Verwendung der neuesten Big-Data-Technologien implementieren.

Modulbezeichnung: Big Data Project

Inhalte des Moduls

Die Studierenden können jedes Thema ihres Interesses, jeden Datensatz oder jede vorhandene Codebasis nach Abstimmung mit dem Dozierenden frei wählen. Die Implementierung eines typischen Big Data-Projekts könnte die folgenden Schritte erfordern:

- Wählen Sie ein zu lösendes Problem aus, z.B. analysieren Sie die Popularität von Filmen in den letzten zehn Jahren und vergleichen Sie die Unterschiede zwischen Brasilien, Frankreich und den USA.
- Wählen Sie die Datensätze aus, z.B. verwenden Sie den Inhalt aus der Internet-Filmdatenbank (IMDB), die in einer relationalen Datenbank gespeichert ist. Reichern Sie die Informationen über Filme mit Dokumenten an, die im Internet gefunden wurden.
- Wählen Sie ein Basissystem mit traditioneller Technologie, z.B. verwenden Sie PostgreSQL, um die in der IMDB gespeicherten Informationen zu analysieren, oder verwenden Sie Ihr bevorzugtes Information Retrieval System, um die Textdokumente zu den Filmen zu analysieren.
- Wählen Sie ein hochmodernes Big-Data-System zum Vergleich mit den Baselines.
- Implementieren Sie die Anwendung sowohl mit dem Basisliniensystem als auch mit dem Big-Data-System.
- Analysieren Sie den Leistungsunterschied der beiden Systeme anhand kleiner Datenmengen.
- Erhöhen Sie den Umfang der Daten beträchtlich und untersuchen Sie die Auswirkungen auf die Leistung.

Durch die Implementierung des Big-Data-Projekts lernen die Studenten die folgenden Aspekte kennen:

Funktionsweise von Big-Data-Systemen:

- - Welche Art von Problemen kann ich mit Big-Data-Systemen lösen?
- - Welche Probleme eignen sich nicht für Big-Data-Systeme?
- - Welche typischen datenwissenschaftlichen Algorithmen werden von Big-Data-Systemen unterstützt?

Performance-Aspekte von Systemen mit Großen Daten:

- - Wie muss ich meine Anwendung neu schreiben, wenn die Größe des Datensatzes um den Faktor 10, 100, 1000 usw. zunimmt?
- - Welche Auswirkungen hat es auf die Performance, wenn die Anzahl der Benutzer um den Faktor 10, 100, 1000 usw. zunimmt?
- - Wie kann ich die Antwortzeit konstant halten?
- - Welche Art von Optimierungsschritten sind für die Implementierung einer Lösung im Unternehmensmaßstab erforderlich?

Benutzerfreundlichkeit von Big Data-Systemen:

- - Wie sieht die Lernkurve der Big-Data-Technologie im Vergleich zur traditionellen Technologie aus?
- - Lohnt es sich für ein kleines, mittleres oder großes Unternehmen, in die Big-Data-Technologie zu investieren, wenn die Studierenden eine spezifische Anwendung implementiert haben?

Vorkenntnisse

https://gpmpublic.zhaw.ch/GPMDocProdDPublic/2_Studium/2_02_Grundlagen_Studium/T_C_L_Modulauspraegungen_SM2025.pdf

Modulbezeichnung: Big Data Project

Lernziele (Kompetenzen)	Die Studierenden...		Kompetenzen	Taxonomiestufen	
	lernen, ihre erstellten Prototypen experimentell zu bewerten und mit einer ausgewählten Basislinie zu vergleichen.		M, F	K3, K6	
	sind in der Lage, über ihr Vorwissen hinauszugehen und geeignete Technologien zu wählen, die den Problemen ihres Kursprojekts entsprechen.		SE	K3	
	sind in der Lage, einen kompletten Projektlebenszyklus im Team zu durchlaufen, vom Projektvorschlag bis zur Projektdurchführung und Präsentation.		SO, M	K3	
	vertiefen die Methoden und Tools aus anderen Kursen (z.B., Data Engineering 1 und Data Engineering 2) durch Anwendung in einem grösseren Kursprojekt.		M, F	K3, K4	
	sind in der Lage, Python und das dazugehörige Data-Science-Ökosystem (z.B. pandas, numpy, scikit-learn) zu nutzen und selbständig in ihrem Kursprojekt anzuwenden.		F, M	K3	
	lernen über die praktische Anwendbarkeit von Big Data Systemen (z.B., Spark) und deren Vor- und Nachteile.		F, M	K3, K4	
Leistungsnachweis	Modulendprüfung	Bewertung	Dauer (Min.)	Gewichtung	Form
	andere	Note		100	gem. Modulvereinbarung
	Leistungsnachweise während dem Semester		Bewertung	Dauer (Min.)	Gewichtung
-		-	-	-	-
Präsenzverpflichtung im Kontaktstudium	Keine				
Lernmaterialien					
Bemerkungen					