

Module description: Big Data Project	
Module Code	t.BA.DS.PM4.20HS
ECTS Credits	4
Language of Instruction/Examination	German
Organizational Unit	InIT
Module Coordinator	Jonathan Fürst
Legal Framework	The module description is part of the legal basis in addition to the general academic regulations. It is binding. During the first week of the semester a written and communicated supplement can specify the module description in more detail.
Module Characteristic	Type 4* 4 lab lessons per semester week and half-class
Module Description	Students gain practical experience of working with Big Data problems. Based on the theoretical foundations of "Data Engineering 1" and "Data Engineering 2", students analyse selected topics from these foundation courses and implement scalable applications using the latest Big Data technologies.
Module Content	<p>Students are free to choose any topic of their interest, any dataset or any existing code base after consultation with the lecturer. Implementing a typical Big Data project could require the following steps:</p> <ul style="list-style-type: none"> • Select a problem to solve, e.g. analyze the popularity of movies over the last ten years and compare the differences between Brazil, France and the USA. • Select the datasets, e.g. use the content from the internet movie database (IMDB) stored in a relational database. Enrich the information about movies with documents found on the internet. • Select a baseline system using traditional technology, e.g. use PostgreSQL to analyze the information stored in IMDB or use your favorite information retrieval system to analyze the text documents about the movies. • Select a state-of-the-art Big Data system to compare against the baselines. • Implement the application using the baselines system as well as the Big Data system. • Analyze the performance difference of both systems using small amounts of data. • Significantly increase the size of the data and study the performance impact. <p>By implementing the Big Data project, students will learn the following aspects:</p> <p>Functionality of Big Data systems:</p> <ul style="list-style-type: none"> • - What kind of problems can I solve with Big Data systems? • - Which problems are not suited for Big Data systems? • - What typical data science algorithms are supported by Big Data systems? <p>Performance aspects of Big Data systems:</p> <ul style="list-style-type: none"> • - How do I need to re-write my application when the size of the dataset increases by a factor of 10, 100, 1000, etc.? • - What is the impact on the performance, when the number of users increases by a factor of 10, 100, 1000, etc.? • - How can I keep the response time constant? • - What kind of optimization steps are required to implement an enterprise-scale solution? <p>Usability of Big Data systems:</p> <ul style="list-style-type: none"> • - What is the learning curve of Big Data technology compared to traditional technology? • - Given a specific use that the students have implemented, does it pay off for a small, medium or large company to invest in Big Data technology?
Prerequisite Knowledge	https://gpmpublic.zhaw.ch/GPMDocProdDPublic/2_Studium/2_02_Grundlagen_Studium/T_C_L_Modulauspraegungen_SM2025.pdf

Module description: Big Data Project

Learning Objectives (Competences)	Students...		Competencies	Taxonomies	
	learn to perform experimental evaluation for their created prototypes, including a comparison with a chosen baseline.		M, F	K3, K6	
	are able to go beyond their prior knowledge and decide on appropriate technologies that match the problems found in their course project.		SE	K3	
	are able to go through a complete project lifecycle in a team, from project proposal to project execution and presentation.		SO, M	K3	
	deepen the methods and tools learned from other courses (e.g., Data Engineering 1 and Data Engineering 2) by applying them in a larger course project.		M, F	K3, K4	
	are able to use Python and its data science ecosystem (e.g., pandas, numpy, scikit-learn) and independently apply it in their course project.		F, M	K3	
	learn about the practical applicability of Big Data Systems (e.g., Spark) with their advantages and disadvantages.		F, M	K3, K4	
Performance Assessment	End-of-module exam	Assessment	Length (min.)	Weighting	Form
	other	Grade		100	acc. to module agreement
	Performance assessment during the semester		Assessment	Length (min.)	Weighting
		-	-	-	-
Classroom Attendance Requirement	None				
Learning material					
Comments					