Courses in English
Course Description

Department	07 Computer Science and Mathematics
Course title	**Big Data Analytics**
Hours per week (SWS)	4
Number of ECTS credits	5
Course objective	The students will be able to understand, describe and apply principles of storing, indexing, and analysing huge amounts of data (big data) in cluster environments. Further, they learn the foundations of programming in Scala and the compute framework Apache Spark. Moreover, they will be able to analyse, conceptualise, implement and evaluate solutions of big data problem.
Prerequisites	advanced programming skills, basic knowledge of data networks
Recommended reading	White, Tom (2017). Hadoop: The Definitive Guide. O'Reilly and Associates. Chambers, Bill & Zaharu, Matei (2018). Spark: The Definitive Guide: Big data processing made simple. O'Reilly UK Ltd. Wills, Josh & Laserson, Uri & Owen, Sean & Ryza, Sandy (2017). Advanced Analytics with Spark: Patterns for Learning from Data at Scale. O'Reilly UK Ltd. Gormley, Clinton & Tong, Zachary (2015). Elasticsearch: The Definitive Guide. O'Reilly and Associates. Schwartz, Jason (2014). Learning Scala: Practical Functional Programming for the JVM. O'Reilly and Associates.
Teaching methods	beamer, whiteboard, Jupyter/Zeppelin notebooks, dashboards (ElasticSearch, Kibana)
Assessment methods	IG 2010: student assignment (40%) + written exam 90min (60%) IG 2019: bonus (30%) + written exam 90min or oral exam
Language of instruction	English
Name of lecturer	Prof. Dr. David Spieler
Email	david.spieler@hm.edu
Link	https://w3-o.cs.hm.edu:8000/public/module/331/
Course content	Many applications of machine learning are based on huge amounts of data which can not be handled on single machines due to the enormous amount of storage and compute demands. The underlying principle of methods and technologies in big data is to distribute storage and computation to clusters of computers. In this course, we will discuss distributed file systems, data sets and computation on the basis of Apache Hadoop (HDFS, YARN, MapReduce) and Apache Spark. Further, a basic introduction to the JVM-based functional programming language Scala will be given followed by topics like data preparation for efficient processing, programming frameworks like MapReduce and Spark, realtime analysis with indexing (ElasticSearch) and dashboard visualisations (Kibana), as well as the distributed implementation of machine learning algorithms.
Remarks	