Module MA-INF 4212	**Data Science and Big Data**		
Workload 180 h	**Credit points** 6 CP	**Duration** 1 semester	**Frequency** every year
Module coordinator	Prof. Dr. Stefan Wrobel		
Lecturer(s)	Dr. Tamas Horvath, PD Dr. Michael Mock		
Classification	**Programme** M. Sc. Computer Science	**Mode** Optional	**Semester** 3. or 4.
Technical skills	Participants acquire in-depth knowledge of different aspects of big data analytics and systems, including distributed processing systems and big data databases, as well as algorithmic techniques for analyzing structured and unstructured data that cannot be stored in a single computer because it has enormous size and/or continuously arrives with such a high rate that requires immediate processing.		
Soft skills	Communicative skills (oral and written presentation of solutions, discussions in teams), self-competences (ability to accept and formulate criticism, ability to analyse, creativity in the context of an "open end" task), social skills (effective team work and project planning).		
Contents	The module is offered every year, each time concentrating on one or more specific issues, such as - architectures and procols for big data systems, - distributed batch and stream processing systems, - non-standard databases for big data, - databases for structured data, - similarity search, - synopses for massive data, - classical data mining tasks for massive data and/or data streams, - mining massive graphs, - applications.		
Prerequisites	**Recommended:** all of the following: MA-INF 4111 – Intelligent Learning and Analysis Systems: Machine Learning MA-INF 4112 – Intelligent Learning and Analysis Systems: Data Mining and Knowledge Discovery		

Format	**Teaching format**	**Group size**	**h/week**	**Workload[h]**	**CP**
	Lecture	60	2	30 T / 45 S	2.5
	Exercises	30	2	30 T / 75 S	3.5
	T = face-to-face teaching; S = independent study				

Exam achievements	Written exam (graded)
Study achievements	Successful exercise participation (not graded)
Forms of media	lectures, exercises, software systems
Literature	- N. Marz and J. Warren: Big Data. Principles and best practices of scalable realtime data systems. Manning Pubn, 2014. - T. White: Hadoop The Definitive Guide. O'REILLY, 2012. - A. Rajaraman and J.D. Ullman.: Mining of Massive Datasets. Cambridge University Press, 2011. - G. Cormode, M. Garofalakis, P.J. Haas, and C. Jermaine: Synopses for Massive Data: Samples, Histograms, Wavelets, Sketches. Foundations and Trends in Databases 4(1-3): 1-294 (2012).