

Georg-August-Universität Göttingen Module M.Inf.1236: High-Performance Data Analytics	6 C 4 WLH
Learning outcome, core skills: Successfully completing the module, students understand <ul style="list-style-type: none"> • the motivation and use-case for large-scale data analytics • performance implications of hardware and software system for large-scale data workloads • the usage of industry-standard tools to solve data analytics problems • algorithms, data structures, data models, tools, and infrastructure for efficient processing of data 	Workload: Attendance time: 56 h Self-study time: 124 h
Course: M.Inf.1236.Lec High-Performance Data Analytics (Lecture, Exercise) <i>Contents:</i> Data-driven science requires the handling of large volumes of data in a quick period of time. Executing efficient workflows is challenging for users but also for systems. This module introduces concepts, principles, tools, system architectures, techniques, and algorithms toward large-scale data analytics using distributed and parallel computing. We will investigate the state-of-the-art of processing data of workloads using solutions in High-Performance Computing and Big Data Analytics. <i>Topics cover:</i> <ul style="list-style-type: none"> • Challenges in high-performance data analytics • Use-cases for large-scale data analytics • Performance models for parallel systems and workload execution • Data models to organize data and (No)SQL solutions for data management • Industry relevant processing models with tools like Hadoop, Spark, and Paraview • System architectures for processing large data volumes • Relevant algorithms and data structures • Visual Analytics • Parallel and distributed file systems Guest talks from academia and industry will be incorporated in teaching that demonstrates the applicability of this topic. Weekly laboratory practicals and tutorials will guide students to learn the concepts and tools. In the process of learning, students will form a learning community and integrate peer learning into the practicals. Students will have opportunities to present their solutions to the challenging tasks in the class. Students will develop presentation skills and gain confidence in the topics.	4 WLH
Examination: Written exam (90 min) or oral exam (approx. 30 min) M.Inf.1236.Mp: High-Performance Data Analytics Examination requirements: <ul style="list-style-type: none"> • Challenges in high-performance data analytics • Use-cases for large-scale data analytics • Performance models for parallel systems and workload execution • Data models to organize data and (No)SQL solutions for data management 	6 C

<ul style="list-style-type: none"> • Industry relevant processing models with tools like Hadoop, Spark, and Paraview • System architectures for processing large data volumes • Relevant algorithms and data structures • Visual Analytics • Parallel and distributed file systems 	
Admission requirements: none	Recommended previous knowledge: Basic programming skills, Basic knowledge of Linux operating systems, Python
Language: English	Person responsible for module: Prof. Dr. Julian Kunkel
Course frequency: each winter semester	Duration: 1 semester[s]
Number of repeat examinations permitted: twice	Recommended semester: Bachelor: 5 - 6; Master: 1 - 4
Maximum number of students: 50	