| Georg-August-Universität Göttingen<br><br>**Module B.Inf.1231: Infrastructures of Data Science** | 6 C<br>4 WLH |
|---|---|

| **Learning outcome, core skills:**<br>Upon completion the course, students<br><br>• understand the basic functions of data science infrastructures and their significance.<br>• understand basic data types and their specifics.<br>• understand the most important technical infrastructures for storing and processing data locally and in the cloud as well as their advantages and disadvantages in relation to data science applications.<br>• can apply the concept of the data lake to basic data science problems.<br>• are able to apply the different steps of data pre-processing to selected data sets.<br>• can identify the characteristics of time series and graph data and are able to recall the functions of DBMSs designed for their processing.<br>• can present the basic tasks of data analysis platforms and can describe them using examples.<br>• can apply methods and tools for the presentation and visualisation of data.<br>• can model basic data science workflows and are able to transfer their knowledge to basic data science projects. | **Workload:**<br>Attendance time:<br>56 h<br>Self-study time:<br>124 h |
|---|---|

| **Course: Infrastructures of Data Science** (Lecture, Exercise)<br>*Contents*:<br>• Data types and their characteristics<br>• Common functions of data science infrastructures<br>• Storage, compute, and cloud infrastructures for data science<br>• Concept of a data lake<br>• Data pre-processing methods and selected tools<br>• Time series and graph data, the respective DBMS, and query languages<br>• Data analytics platforms<br>• Data presentation and visualization<br>• Data science workflows and selected infrastructure components | 4 WLH |
|---|---|
| **Examination: In-class, written exam (90 min) or oral exam (approx. 30 min.)**<br>**Examination prerequisites:**<br>Students complete 50% of the homework exercises.<br>**Examination requirements:**<br>Through the examination students demonstrate that they are able to describe basic functions of (cloud-based) data science infrastructures as well as to specify and identify basic data types. Students can also prove their understanding of data lakes and can apply their knowledge of MapReduce and Hadoop in that particular context. They can analyse basic data pre-processing problems and sketch common solutions. Student can show that they understand time series and graph data as well as the corresponding DBMS and that they can present common tasks of data analysis platforms. Through the examination, students also demonstrate their ability to select appropriate methods for visualising data and show that they are able to create basic data science workflows. | 6 C |

| Admission requirements:<br>none | Recommended previous knowledge:<br>Python and basic database knowledge<br>(recommended, not mandatory) |
|---|---|
| Language:<br>English | Person responsible for module:<br>Hon.-Prof. Dr. Philipp Wieder |
| Course frequency:<br>each summer semester | Duration:<br>1 semester[s] |
| Number of repeat examinations permitted:<br>twice | Recommended semester:<br>Bachelor: 3 - 6; Master: 1 - 2 |
| Maximum number of students:<br>50 | |