Data Engineering

**Course Description**

The focus of this first course in the Data Engineering elective module is to introduce students to important principles, concepts, methods and approaches in this subject domain.

In order to achieve this goal, the course moves from an exposition of the foundational principles of data engineering to a thorough treatment of the core data processing classes. Modern architectural paradigms such as Microservices are explained, and important factors in data governance and protection are addressed. In this context, students are enabled to reflect on modern data protection principles and their societal implications and implement these principles into large-scale data-intensive systems. Aspects of cloud computing are introduced via an overview of the most common offerings on the market. Finally, a state-of-the-art agile perspective on the operation of data pipelines is given by an exposition to the emerging notion of DataOps and the productivity tools around it to facilitate working in interdisciplinary teams.

**Contents**

1. Foundations of Data Systems
   1. Reliability
   2. Scalability
   3. Maintainability
2. Data Processing at Scale
   1. Batch Processing
   2. Stream and Complex Event Processing
3. Microservices
   1. Introduction to Microservices
   2. Implementing Microservices
4. Governance & Security
   1. Data Protection
   2. Data Security
   3. Data Governance
5. Common Cloud Platforms & Services
   1. Amazon AWS
   2. Google Cloud
   3. Microsoft Azure
6. Data Ops
   1. Defining Principles
   2. Containerization
   3. Building Data Pipelines