

Course: Programming 6

credits: 5

Course code	BFVM25PROGRAM6	Modes of delivery	Assignment
Name	Programming 6		Lecture
Study year	2026-2027		Tutorial
ECTS credits	5	Assessments	Programming 6 - Assignment
Language	English		
Coordinator	F. Feenstra		

Learning outcomes

- You design, implement, and optimize scalable data processing and or ML pipelines using Snakemake, ensuring efficient resource usage, high performance, and reproducibility for practical big data applications.
- You create a comprehensive model card for an ML model that documents its intended use, data provenance, performance metrics, limitations, and ethical considerations, and you critically evaluate how this artifact supports transparency, responsible deployment, and effective stakeholder communication throughout the ML lifecycle.
- You design and implement an automated documentation system, integrated into a pipeline, that systematically captures installation procedures, usage instructions, and model versioning, and you critically justify how this documentation ensures traceability, reproducibility, and auditability of ML models across the entire operational lifecycle.

Content

This course builds on foundational concepts of distributed computing frameworks, such as Spark, job schedulers like SLURM, and distributed storage and batch processing systems like Hadoop.

In this course, you will explore additional essential tools for big data and distributed computing, such as Snakemake. You will gain hands-on experience in using these tools to design and implement efficient, scalable solutions for big data and high-performance computing challenges. The course will also cover best practices for integrating machine learning into operational workflows (ML-OPS), ensuring robust and optimized performance.

To enroll in and be evaluated for this subject, completion and or granted exemptions of 3 preparatory courses is required.

Included in programme(s)

Master Data Sciences for the Life Science

School(s)

Engineering, Life Sciences & ICT