Vak: Programming 5 — credits: 5

Vakcode	BFVM25PROGRAM5
Naam	Programming 5
Studiejaar	2025-2026
ECTS credits	5
Taal	Engels
Coördinator	-

Werkvormen	Hoorcollege Opdracht Werkcollege
Toetsen	Programming 5 - Opdracht

Leeruitkomsten

Following this course you will;

- Understand the principles of parallel programming on large clusters.
- Be able to use the Apache Spark framework to process large datasets, larger than fit on one computer.
- Understand the performance of your program, and structure data and operations in such a way that you maximize throughput.

Inhoud

This course introduces you to parallel programming on large clusters, using the Apache Spark framework. Spark is a good example of a big-data processing framework, which gives you a set of tools to operate on very large datasets on multiple computers transparently. In other words, you don't program any differently than you would, using the same abstraction (dataframes) on a single node; the framework takes care of all network programming overhead.
There are a couple of extra tools that we will use, which are not directly part of the Spark eco-system, but which are both essential knowledge for Data Scientists, and which will help you in your work with Spark.
These are:

- Structured Query Language (SQL), the standard programming language for interacting with relational databases.
- SLURM, a workload manager, which allows you to run programs on a cluster of computers in a controlled way.
- Cassandra, a distributed database, which is a good example of a NoSQL database, and which is often used in combination with Spark.

This is a 7-week practical research course, meaning you will be graded on programs that you submit, and that those programs are dealing with biological or biomedical research data. The course is taught in Python, and you are expected to have a good understanding of Python programming before you start this course; otherwise you will not be able to complete it.

Opgenomen in opleiding(en)

Master Data Sciences for the Life Science - 4 semesters

School(s)

Instituut voor Life Science & Technology

share your talent. move the world.

De ECTS onderwijscatalogus van de Hanzehogeschool Groningen wordt met de grootst mogelijke zorg samengesteld. Het is echter mogelijk dat de inhoud van de catalogus -en de daarin vervatte informatie- verouderd, incompleet of onjuist is. Aan de inhoud van de catalogus kunnen dan ook geen rechten worden ontleend.