

Vak: Big data computing

credits: 3

Vakcode BFBVH4BDC
Naam Big data computing
Studiejaar 2019-2020
ECTS credits 3
Taal Nederlands
Coördinator M. Herber

Werkvormen Werkcollege
Toetsen TOETS-01 - Opdracht

Leeruitkomsten

- Kennis opdoen van bestaande grid-computing systemen.
- Leren programmeren met de Hadoop API (Java/Python).
- Leren hoe bestaande programma's en algoritmen aan te passen voor gebruik met Hadoop.

Inhoud

Bioinformatici moeten bijzonder grote hoeveelheden data kunnen verwerken; gigabytes zo niet terabytes. Individuele computers zijn hier vaak niet krachtig genoeg voor.

Een oplossing is dan om meerdere computers aan het rekenen te zetten.

Hoe dit te organiseren is een zich continu ontwikkelend veld. In dit vak komt de theorie van een aantal verschillende systemen aan bod (OpenMPI, Condor, SLURM, Hadoop, Spark), en wordt de focus gelegd op het Hadoop systeem.

Dit Grid-computing systeem is erg populair in de "Big Data" wereld maar vergt wel aanpassing van bestaande programma's naar een bepaald format: het Map/Reduce patroon. Aan de hand van de Weka toolkit uit de module Advanced Datamining worden een aantal algoritmen uit Thema 11 aangepast voor Hadoop.

De geleerde technieken zullen verder toegepast worden in het project van Thema 12 "Big Data & Machine Learning". Online documentatie op <http://hadoop.apache.org/> en andere webstes. Optioneel boek "Hadoop: the Definitive Guide" van Tom White (O'Reilly, 3rd Edition of later).

Blackboard course thema 12

Opgenomen in opleiding(en)

Minor High Throughput/High-Performance Biocomputing

School(s)

Instituut voor Life Science & Technology

share your talent. move the world.