UNIVERSITY OF GOTHENBURG

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DIT066 Computational techniques for large-scale data, 7.5 credits

Beräkningsmetoder för storskaliga data, 7,5 högskolepoäng

Second Cycle

Confirmation

This course syllabus was confirmed by Department of Computer Science and Engineering on 2023-11-14 to be valid from 2025-01-20, spring semester of 2025.

Field of education: Science 100%
Department: Department of Computer Science and Engineering

Position in the educational system

The course can be part of the following programmes: 1) Computer Science, Master's Programme (N2COS), 2) Applied Data Science Master's Programme (N2ADS) and 3) Software Engineering and Management Master's Programme (N2SOF)

Main field of studies	*Specialization*
Data Science	A1N, Second cycle, has only first-cycle course/s as entry requirements
Computer Science	A1N, Second cycle, has only first-cycle course/s as entry requirements

Entry requirements

To be eligible to the course, the student should have a Bachelor's degree in any subject, or have successfully completed 90 credits of studies in computer science, software engineering, or equivalent. Specifically, at least 15 credits of successfully completed courses in programming, of which at least 7.5 credits in Python programming, or equivalent are required. The student needs to have successfully completed a course in probability theory or statistics, for example DIT862 or MSG810 or similar.

Applicants must prove knowledge of English: English 6/English B or the equivalent level of an internationally recognized test, for example TOEFL, IELTS.

Learning outcomes

On successful completion of the course the student will be able to:

Knowledge and understanding

- discuss important technological aspects when designing and implementing analysis solutions for large-scale data,
- explain differences between parallel programming models,
- describe data structures and algorithms for big data and discuss their utility.

Competence and skills

- implement applications for transforming and analyzing large-scale data with different parallel software frameworks,
- use algorithms and datastructures for computations with large-scale data.

Judgement and approach

- suggest appropriate computational infrastructures and methodological approaches for analysis tasks and discuss their advantages and drawbacks,
- discuss advantages and drawbacks of different strategies of parallelization,
- decide between algorithmic and parallelization-based approaches for accelerating computational workloads.

Course content

The aim of this course is to deepen the students' knowledge and skills and familiarize them with the technical and technological side of data science, including software respectively hardware environments. The course will introduce aspects of designing and implementing large-scale data science solutions.

In particular, the course will include:

- an overview of computer architectures, algorithmic approaches, and high-performance computing infrastructures with a focus on limitations for processing large-scale data,
- an introduction to relevant frameworks for cluster computing with large-scale data,
- implementation of data analysis tools on a cluster using Python and appropriate software frameworks,
- data structures and algorithms, such as index structures, which can greatly accelerate computations with large-scale data

Sub-courses

1. **Written hall examination** *(Skriftlig salstentamen),* 3 credits
   Grading scale: Pass with distinction (5), Pass with credit (4), Pass (3) and Fail (U)
2. **Assignments** *(Inlämningsuppgifter),* 4.5 credits

Grading scale: Pass with distinction (5), Pass with credit (4), Pass (3) and Fail (U)

Form of teaching

Lectures, computer lab sessions, and exercise sessions.

Language of instruction: English

Assessment

The course is examined by a written hall examination, as well as mandatory written assignments, some of which will be carried out individually and others will be carried out in groups of normally 2-4 students. There will be non-obligatory individual assignments which grant bonus points for the written examination. These bonus points are valid for the two next scheduled re- examinations.

A student who has taken two exams in a course or part of a course without obtaining a pass grade is entitled to the nomination of another examiner. The student needs to contact the department for a new examiner, preferably in writing, and this should be approved by the department unless there are special reasons to the contrary (Chapter 6 Section 22 of the Higher Education Ordinance).

If a student has received a recommendation from the University of Gothenburg for special educational support, where it is compatible with the learning outcomes of the course and provided that no unreasonable resources are required, the examiner may decide to allow the student to sit an adjusted exam or alternative form of assessment.

In the event that a course has ceased or undergone major changes, students are to be guaranteed at least three examination sessions (including the ordinary examination session) over a period of at least one year, but no more than two years, after the course has ceased/been changed. The same applies to placements and professional placements (VFU), although this is restricted to just one additional examination session.

Grades

The grading scale comprises: Pass with distinction (5), Pass with credit (4), Pass (3) and Fail (U).
To pass the course, all mandatory components must be passed. To earn a higher grade than Pass, a higher weighted average from the grades of the components is required.

Course evaluation

The course is evaluated through meetings both during and after the course between teachers and student representatives. Further, an anonymous questionnaire is used to ensure written information. The outcome of the evaluations serves to improve the course by indication which parts could be added, improved, changed or removed.

Additional information

The course is a joint course together with Chalmers.

Course literature to be announced the latest 8 weeks prior to the start of the course.

The course replaces the course DIT065, 7.5 credits. The course cannot be included in a degree which contains DIT065. Neither can the course be included in a degree which is based on another degree in which the course DIT065 is included.