UNIVERSITY OF GOTHENBURG

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

DIT872 Techniques for Large-scale Data, 7.5 credits

Tekniker för storskaliga data, 7,5 högskolepoäng

Second Cycle

Confirmation

This course syllabus was confirmed by Department of Computer Science and Engineering on 2017-12-19 and was last revised on 2018-02-09 to be valid from 2018-08-19, autumn semester of 2018.

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

Position in the educational system

The course is compulsory within the programme Applied Data Science Master's Programme. It is also a single subject course at the University of Gothenburg.

The course can be part of the following programmes: 1) Computer Science, Master's Programme (N2COS), 2) Applied Data Science Master's Programme (N2ADS), 3) Computer Science, Bachelor´s Programme (N1COS) and 4) Software Engineering and Management Master's Programme (N2SOF)

Main field of studies	*Specialization*
Computer Science	A1N, Second cycle, has only first-cycle course/s as entry requirements
Software Engineering	A1N, Second cycle, has only first-cycle course/s as entry requirements
Data 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 are required, as well as a database course of at least 7.5 credits (e.g. DIT621 Databases or DIT032 Data Management, or equivalent).

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

After completion of the course the student should be able to:

Knowledge and understanding

- discuss important technological aspects when designing and implementing analysis solutions for large-scale data,
- describe data models and software standards for sharing data on the web.

Competence and skills

- implement applications for transforming and analyzing large-scale data with appropriate software frameworks,
- provide access and utilize structured data over the web with appropriate data models and software tools.

Judgement and approach

- suggest appropriate computational infrastructures for analysis tasks and discuss their advantages and drawbacks,
- discuss advantages and drawbacks of different strategies for dissemination of data,
- discuss large-scale data processing from an ethical point of view.

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 relevant data models, and 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 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,
- an overview of non-relational database technologies,
- semantic web and related technologies,

- an overview of ethical questions regarding large-scale data, e.g. with respect to licenses, accessibility, and anonymisation.

Sub-courses

1. **Written examination** *(Skriftlig tentamen),* 4 higher education credits
   Grading scale: Pass with Distinction (VG), Pass (G) and Fail (U)
2. **Assignments** *(Inlämningsuppgifter),* 3.5 higher education credits
   Grading scale: Pass with Distinction (VG), Pass (G) and Fail (U)

Form of teaching

Lectures, computer lab sessions, and exercise sessions.

Language of instruction: English

Assessment

The course is examined by an individual written exam carried out in an examination hall, 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.

If a student, who has failed the same examined component twice, wishes to change examiner before the next examination, a written application shall be sent to the department responsible for the course and shall be granted unless there are special reasons to the contrary (Chapter 6, Section 22 of Higher Education Ordinance).

In cases where a course has been discontinued or has undergone major changes, the student shall normally be guaranteed at least three examination occasions (including the ordinary examination) during a period of at least one year from the last time the course was given.

Grades

The grading scale comprises: Pass with Distinction (VG), Pass (G) and Fail (U).
A Pass grade (G) for the entire course requires at least a Pass grade for all sub-courses.

To be awarded Pass with Distinction (VG) for the full course, the grade VG must be obtained on both the sub-courses.

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 indicating 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 DIT871, 7.5 credits. The course cannot be included in a degree which contains DIT871. Neither can the course be included in a degree which is based on another degree in which the course DIT871 is included.