



# ID2221 Data-Intensive Computing 7.5 credits

Data-intensiv databehandling

This is a translation of the Swedish, legally binding, course syllabus.

If the course is discontinued, students may request to be examined during the following two academic years

## Establishment

Course syllabus for ID2221 valid from Autumn 2016

## Grading scale

A, B, C, D, E, FX, F

## Education cycle

Second cycle

## Main field of study

Computer Science and Engineering

## Specific prerequisites

## Language of instruction

The language of instruction is specified in the course offering information in the course catalogue.

# Intended learning outcomes

The course complements distributed systems courses, with a focus on processing, storing and analyzing massive data. It prepares the students for master projects, and Ph.D. studies in the area of data-intensive computing systems. The main objective of this course is to provide the students with a solid foundation for understanding large scale distributed systems used for storing and processing massive data.

More specifically after the course is completed the student will be able to

- explain the architecture and properties of the computer systems needed to store, search and index large volumes of data
- describe the different computational models for processing large data sets for data at rest (batch processing) and data in motion (stream processing)
- use various computational engines to design and implements nontrivial analytics on massive data
- explain the different models for scheduling and resource allocation computational tasks on large computing clusters
- elaborate on the tradeoffs when designing efficient algorithms for processing massive data in a distributed computing setting.

## Course contents

Topics:

- Distributed file systems
- No SQL databases
- Scalable messaging systems
- Big Data execution engines: Map-Reduce, Spark
- High level queries and interactive processing: Hive and Spark SQL
- Stream processing
- Graph processing
- Scalable machine learning
- Resource management.

## Examination

- LAB1 - Programming Assignments, 3.0 credits, grading scale: P, F
- TEN1 - Examination, 4.5 credits, grading scale: A, B, C, D, E, FX, F

Based on recommendation from KTH's coordinator for disabilities, the examiner will decide how to adapt an examination for students with documented disability.

The examiner may apply another examination format when re-examining individual students.

Written examination. Laboratory tasks.

## **Ethical approach**

- All members of a group are responsible for the group's work.
- In any assessment, every student shall honestly disclose any help received and sources used.
- In an oral assessment, every student shall be able to present and answer questions about the entire assignment and solution.