



INSTITUTIONEN FÖR DATA- OCH INFORMATIONSTEKNIK

DIT346 Tekniker för storskalig data, 7,5 högskolepoäng

Techniques for Large-scale Data, 7.5 credits

Avancerad nivå / Second Cycle

Fastställande

Kursplanen är fastställd av Institutionen för data- och informationsteknik 2019-12-02 att gälla från och med 2021-01-18, vårterminen 2021.

Utbildningsområde: Naturvetenskapligt 100 %

Ansvarig institution: Institutionen för data- och informationsteknik

Inplacering

Kursen är obligatorisk inom programmet N2ADS. Den ges även som fristående kurs vid Göteborgs universitet.

Kursen kan ingå i följande program: 1) Datavetenskapligt program (N1COS), 2) Computer Science, Master's Programme (N2COS), 3) Software Engineering and Management masterprogram (N2SOF), 4) Applied Data Science masterprogram (N2ADS) och 5) Game Design & Technology masterprogram (N2GDT)

Huvudområde

Datavetenskap

Data Science

Software Engineering

Fördjupning

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

A1N, Avancerad nivå, har endast kurs/er på grundnivå som förkunskapskrav

Förkunskapskrav

För tillträde till kursen krävs att studenten har en examen på kandidatnivå inom något ämne, eller har minst 90 hp i datavetenskap, software engineering eller motsvarande. Specifikt krävs minst 15 hp avklarade kurser i programmering eller motsvarande, samt att studenten genomgått en kurs i databaser på minst 7.5 hp (t.ex. DIT621 Databaser

eller DIT032 Systematiskt datahantering).

Studenten ska även ha avklarat en kurs i sannolikhetssteori eller statistik, exempelvis DIT862 eller MSG810 eller liknande.

Följande kunskapsnivå i Engelska krävs; Engelska 6/Engelska B eller motsvarande från ett erkänt internationellt test, t.ex. TOEFL, IELTS.

Lärandemål

Efter godkänd kurs ska studenten kunna:

Kunskap och förståelse

- diskutera de viktigaste teknologiska aspekterna vid design och implementering av system för analys av storskaliga data,
- diskutera indexstrukturer och nyttan med dem,
- beskriva datamodeller och mjukvarustandarder för internetdelning av data.

Färdigheter och förmåga

- implementera tillämpningar för att transformera och analysera storskaliga data med hjälp av tillämpbara mjukvaruramverk,
- tillgängliggöra och utnyttja strukturerade data över internet med hjälp av lämpliga datamodeller och mjukvaruverktyg.

Värderingsförmåga och förhållningssätt

- föreslå lämpliga tekniska infrastrukturer för dataanalytiska uppgifter och argumentera fördelar och nackdelar med dessa,
- diskutera fördelar och nackdelar med olika strategier för spridning av data,
- diskutera metoder för parallellism och återställning i databassystem,
- diskutera effektiviteten för sökfrågeplaner,
- diskutera storskalig datahantering utifrån ett etiskt perspektiv.

Innehåll

Kursens mål är att fördjupa studenternas kunskaper och färdigheter inom den tekniska sidan av data science, inklusive de relevanta datamodellerna, samt lämpliga mjukvaru- och hårdvarumiljöer. Kursen kommer att introducera aspekter av design och implementation av storskaliga data science-lösningar.

Framför allt kommer kursen att innefatta:

- en överblick av datorarkitekturer, algoritmiska tillvägagångssätt, och högpresterande beräkningsinfrastrukturer, med ett fokus på de begränsningar som

- finns vid behandling av storskaliga data,
- en inledning till de relevanta ramverken för bearbetning av data med hjälp av datorkluster,
 - implementering av dataanalysverktyg på ett kluster med hjälp av Python och lämpliga mjukvaruramverk,
 - indexstrukturer, hantering och optimering av sökfrågor; parallellism, återställning,
 - en överblick över ickerelationella databasteknologier,
 - Semantic Web och besläktade teknologier,
 - en översikt över etiska frågor inom storskalig datahantering.

Delkurser

1. **Skriftlig salstentamen** (*Written hall examination*), 4,5 hp
Betygsskala: Väl godkänd (VG), Godkänd (G) och Underkänd (U)
2. **Inlämningsuppgifter** (*Assignments*), 3 hp
Betygsskala: Godkänd (G) och Underkänd (U)

Former för undervisning

Föreläsningar, datorlaborationer och övningar.

Undervisningsspråk: engelska

Former för bedömning

Kursen examineras genom en individuell skriftlig salstentamen, samt obligatoriska skriftliga inlämningsuppgifter som redovisas genom rapportinlämning, varav vissa kommer att genomföras individuellt och andra i grupper om normalt 2-4 studenter.

Det kommer att ges frivilliga individuella uppgifter som ger bonuspoäng på tentamen. Dessa bonuspoäng är giltiga under de två nästkommande schemalagda omtentamina..

Om student som underkänts två gånger på samma examinerande moment önskar byte av examinator inför nästa examinationstillfälle, ska sådan begäran inlämnas skriftligt till kursansvarig institution och bifallas om det inte finns särskilda skäl däremot (HF 6 kap § 22).

I det fall en kurs har upphört eller genomgått större förändringar ska studenten i normalfallet garanteras tillgång till minst tre provtillfällen (inklusive ordinarie provtillfälle) under en tid av åtminstone ett år med utgångspunkt i kursens tidigare uppläggning.

Betyg

På kursen ges något av betygen Väl godkänd (VG), Godkänd (G) och Underkänd (U). För godkänt betyg på hel kurs krävs godkänt betyg på samtliga delkurser. För att få betyget Väl Godkänd (VG) på hela kursen krävs därutöver betyget VG på delkursen Skriftlig tentamen.

Kursvärdering

Kursen utvärderas genom möten både under och efter kursen mellan lärare och studentrepresentanter. Därutöver används en anonym enkät för att få skriftlig information. Resultatet av utvärderingen används för att förbättra kursen genom att visa på delar som kan läggas till, förbättras, ändras eller tas bort.

Övrigt

Kursen är samläst med Chalmers.

Kurslitteratur kommer att publiceras senast 8 veckor innan kursstart.

Kursen ersätter kursen DIT872, 7,5 hp. Den här kursen kan inte ingå i en examen som innehåller DIT872. Den kan inte heller ingå i en examen som bygger på en annan examen där DIT872 ingår.