



FID3019 Avancerad kurs i data-intensiv databehandling 7,5 hp

Advanced course in Data-Intensive Computing

När kurs inte längre ges har student möjlighet att examineras under ytterligare två läsår.

Fastställande

Kursplan för FID3019 gäller från och med HT17

Betygsskala

Utbildningsnivå

Forskarnivå

Särskild behörighet

Inskrivna som forskarstuderande.

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Kursen kompletterar kurserna i distribuerade system med ett fokus på bearbetning, lagring och analys av massiva datamängder. Den förbereder studenterna på forskarutbildning inom området data-intensive computing systems.

Huvudsyftet med denna kurs är att erbjuda studenten en solid grund för att förstå storskaliga distribuerade system som används för att lagra och bearbeta massiva datamängder.

Mer specifikt kommer studenter efter avslutad kurs att kunna

- förklara arkitekturen och de egenskaper hos datorsystem som behövs för att lagra, utsöka och indexera stora datavolymer
- beskriva olika beräkningsmodeller för att bearbeta stora datamängder för statistiska data (batchbearbetning) och för data i rörelse (stream processing)
- använda olika beräkningssystem för design och implementation av icke-triviala analysmetoder på massiva data
- förklara olika modeller för att schemalägga och allokera resurser för beräkningsuppgifter på stora datorkluster
- förklara avvägningarna som görs vid design av effektiva algoritmer för att bearbeta stora datamängder i en distribuerad beräkning.

Kursinnehåll

Ämnen:

- Distribuerade filsystem
- “No-SQL”-databaser
- Skalbara meddelandesystem
- Exekveringsmaskiner för Big Data: Map-Reduce, Spark
- Högnivå-frågor och interaktiv bearbetning: Hive och Spark SQL
- Stream processing
- Graph processing
- Skalbar maskininlärning
- Resurshantering.

Kursupplägg

Kursen är organiserad som en läskurs. Varje student väljer ett antal artiklar och för varje artikel skall studenten göra följande:

* noggrant läsa och analysera artikeln.

* muntligt presentera artikelns innehåll, inkluderande metod och bidrag, för de övriga kursdeltagarna och kursens exainator. Presentationen inklusive en diskussion bör ta ungefär en timme.

* skriva en kritik av artikeln som täcker speciellt: en sammanfattning av dess bidrag, metod, signifikans, teknisk och experimentell kvalitet, och presentationskvalitet.

Förutom att presentera fyra artiklar vardera, skall studenterna läsa några av de artiklar som getts till de övriga deltagarna, delta i deras presentationer och aktivt bidra till diskussionen om deras artiklar.

Kurslitteratur

Latest papers in the area of Data intensive Computing from high-quality international venues.

Examination

Examinator beslutar, baserat på rekommendation från KTH:s handläggare av stöd till studenter med funktionsnedsättning, om eventuell anpassad examination för studenter med dokumenterad, varaktig funktionsnedsättning.

Examinator får medge annan examinationsform vid omexamination av enstaka studenter.

P/F

Övriga krav för slutbetyg

Kursen betygsätts med skalan P/F (godkänd/icke godkänd), baserat på en godkänd presentation, levererandet av en vetenskapligt sund rapport och identifieringen av lämpliga artiklar för läslistan. Dessförutan måste studenten delta i minst 75% av kursens seminarier

Etiskt förhållningssätt

- Vid grupparbete har alla i gruppen ansvar för gruppens arbete.
- Vid examination ska varje student ärligt redovisa hjälp som erhållits och källor som använts.
- Vid muntlig examination ska varje student kunna redogöra för hela uppgiften och hela lösningen.