



IV2014 Data Warehousing 7,5 hp

Data Warehousing

Fastställande

Kursplan för IV2014 gäller från och med HT10

Betygsskala

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

Utbildningsnivå

Avancerad nivå

Huvudområden

Särskild behörighet

För fristående studenter krävs:

- 180 högskolepoäng (hp) i data- och systemvetenskap/IT/datavetenskap eller motsv
- kunskaper i eng B eller motsvarande

Undervisningsspråk

Undervisningsspråk anges i kurstillfällesinformationen i kurs- och programkatalogen.

Lärandemål

Kursens övergripande mål är att ge kunskaper om och förståelse för en specifik typ av informationssystem, nämligen datalager (eng. Data Warehouses) – deras strategiska roll, användning och nytta i organisationer; arkitektur och underliggande tekniker för realisering av datalager; analys- och implementeringsprocessen.

Studenten förväntas efter kursen kunna:

1. förklara centrala begrepp avseende datalager, särskilt Kimballs terminologi inom området
2. givet en domänbeskrivning utforma och dokumentera dimensionella modeller för ett datalager
3. analysera, utforma och dokumentera modeller och rutiner för dataöverföring från ett antal angivna källor till ett datalager
4. implementera dimensionella modeller i ett givet system, befolka dessa modeller med data och använda front-end -system för att extrahera och analysera de data som finns i datalagret. Använda ett ETL-system för att extrahera data från olika filer och ladda upp det i relationstabeller.
5. sammanfatta, presentera och värdera resultat ur aktuell vetenskaplig litteratur inom kursens område

Kursinnehåll

Följande delar kommer att behandlas i kursen:

- Dimensionell modellering - grunder och principer
- Typiska konstruktioner (mönster) inom dimensionell modellering
- Fysisk design, typiska indexeringstekniker och användningen av aggregat-tabeller
- Utvecklingsprocess av datalager
- ETL (Extract Transform Load) -fasen av utvecklingsprocessen
- Skillnader mellan datalager och traditionella relationsdatabaser

Kursupplägg

Kursen har ett traditionellt upplägg. Undervisningen består av föreläsningar, lektioner och seminarier. Med stöd av kursens föreläsningar och lektioner skall studenterna genomföra fyra inlämningsuppgifter. Dessa omfattar: analys och design av ett datalager; analys och design av ETL process; implementering av en lösning i ett datalager verktyg samt läsning i datalagern via front-end verktyg; samt en teoretisk uppgift. Alla inlämningsuppgifterna, förutom verktygsuppgiften, redovisas på obligatoriska seminarier. Inlämningsuppgifterna genomförs i grupper om fyra.

OBS! Detta är inte en distanskurs.

Kurslitteratur

Preliminärt:

- R. Kimball & M. Ross: The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd Edition (Upplaga: 2), John Wiley & Sons, Inc., 2002, 0-471-20024-7
- Kompendium med artiklar.

Examination

- INL1 - Inlämningsuppgift, 3,0 hp, betygsskala: P, F
- TEN1 - Examination, 4,5 hp, betygsskala: A, B, C, D, E, FX, F

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 enskilda studenter.

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

Examination för kursen består av två delar:

- Tentamen
- Inlämningsuppgifter

Inlämningsuppgifternas del består av fyra inlämningsuppgifter:

- Inlämningsuppgift 1: Multidimensionell modellering
- Inlämningsuppgift 2: ETL
- Inlämningsuppgift 3: Verktysövning
- Inlämningsuppgift 4: Artikelpresentation

De första tre uppgifterna utförs i grupper om 4. Den sista uppgiften utförs individuellt.

BETYGSKRITERIER FÖR MÅLEN₁. Förklara centrala begrepp avseende datalager, särskilt Kimballs terminologi inom området

A Kunna korrekt definiera, exemplifiera, använda och jämföra minst 95% av ett urval centrala termer och begrepp inom datalager området, särskild datalager arkitektur, dimensionell modellering, ETL processen, front-end verktyg

B Kunna korrekt definiera, exemplifiera, använda och jämföra minst 90% av ett urval centrala termer och begrepp inom datalager området, särskild datalager arkitektur, dimensionell modellering, ETL processen, front-end verktyg

C Kunna korrekt definiera, exemplifiera, använda och jämföra minst 85% av ett urval centrala termer och begrepp inom datalager området, särskild datalager arkitektur, dimensionell modellering, ETL processen, front-end verktyg

D Kunna korrekt definiera, exemplifiera, använda och jämföra minst 80% av ett urval centrala termer och begrepp inom datalager området, särskild datalager arkitektur, dimensionell modellering, ETL processen, front-end verktyg

E Kunna korrekt definiera, exemplifiera, använda och jämföra minst 75% av ett urval centrala termer och begrepp inom datalager området, särskild datalager arkitektur, dimensionell modellering, ETL processen, front-end verktyg

Fx Kunna korrekt definiera, exemplifiera, använda och jämföra minst 60% av ett urval centrala termer och begrepp inom datalager området, särskild datalager arkitektur, dimensionell modellering, ETL processen, front-end verktyg

2. Givet en domänbeskrivning utforma och dokumentera dimensionella modeller för datalager.

För att uppnå ett visst betyg skall den studerande kunna utforma en modell som har nedanstående egenskaper i den grad som anges i cellerna.

A B C D E Fx

Syntaktiskt korrekt Högst ett fel Högst ett fel Högst ett fel Högst ett fel Högst ett fel Högst ett fel

Semantiskt korrekt 95% 90% 80% 70% 60% 50%

Fullständig 95% 90% 80% 70% 60% 50%

Icke-redundant 5% 20% 30% 50% 75% 75%

Lättförståelig Helt Nästan Till stor del Till stor del Acceptabel Acceptabel

Syntaktiskt korrekt betyder språkligt och ritningstekniskt korrekt

Semantiskt korrekt betyder innehållsmässigt korrekt, dvs det som finns i modellen stämmer med verkligheten

Fullständig betyder att alla viktiga delar är med

Icke-redundant betyder att inga onödiga delar är med

Krav för A innebär i klartext att modellen skall innehålla högst ett syntaktiskt fel, den skall till minst 95% överensstämma med den verklighet som modelleras, den skall till minst 95% inkludera de viktiga delarna av verkligheten, den skall till högst 5% bestå av irrelevanta delar, och den skall vara helt lättförståelig.

3. Utforma och dokumentera modeller och rutiner för data överförning från ett antal angivna källor till en datalager

Som ovan

4. Implementera dimensionella modeller i ett givet system, fylla dessa med data och använda front-end system för att extrahera och analysera den data som finns i datalagern. Använda ett ETL system för att extrahera data from olika filer och ladda upp det i relations tabeller.

Godkänd Självständigt kunna implementera en given modell i ett givet datalager system, fylla den med data, samt använda två front-end system för att extrahera data som svarar på ett antal givna frågor. Jämföra de två front-end system som användes. Självständigt kunna använda ett givet ETL system för att extrahera data from olika filer och ladda upp det i relations tabeller.

Fx Delvis kunna implementera en given modell i ett datalager system, delvis fylla den med data, samt delvis använda två front-end system för att extrahera data som svarar på ett antal givna frågor. Delvis kunna använda ett givet ETL system för att extrahera data from olika filer och ladda upp det i relations tabeller.

5. Sammanfatta, presentera och värdera resultat ur aktuell vetenskaplig litteratur inom kursens område

Godkänd Självständigt kunna läsa och sätta sig in i en vald artikel, sammanfatta den, förbereda presentationsmaterial, presentera och diskutera den.

Fx Delvis kunna läsa och sätta sig in i en vald forskningsartikel, sammanfatta den, förbereda presentationsmaterial, presentera och diskutera den.

BETYGSKRITERIER FÖR EXAMINATION AV DELAR

Tentamen

Tentamen mäter hur väl mål 1,2 och 3 är uppnådda och betygsätts enligt följande:

A Alla målen (dvs. 1, 2 and 3) uppnås med betyget A

B Alla målen (dvs. 1, 2 and 3) uppnås med betyget B

C Alla målen (dvs. 1, 2 and 3) uppnås med betyget C

D Alla målen (dvs. 1, 2 and 3) uppnås med betyget D

E Alla målen (dvs. 1, 2 and 3) uppnås med betyget E

Fx Alla målen (dvs. 1, 2 and 3) uppnås med betyget Fx

Inlämningsuppgifter

Inlämningsuppgift 1 mäter mål 2

Inlämningsuppgift 2 mäter mål 3

Inlämningsuppgift 3 mäter mål 4

Inlämningsuppgift 4 mäter mål 5

Inlämningsuppgifterna (tillsammans) betygsätts enligt följande:

Godkänd: mål 2 och 3 skall uppnås med minst betyget C and mål 4 and 5 skall uppnås med betyget Godkänd.

Fx mål 2, 3, 4 och 5 skall uppnås med minst betyget Fx

Studenter som missar den skriftliga tentamen kommer att ges möjlighet till en skriftlig omtenta under samma läsår. Omtentan håller samma kvalitet som tentamen. Studenter som missar omtentan hänvisas till den skriftliga examinationen av delkursen nästföljande år. Studenter som misslyckas med eller missar deadline till inlämningsuppgifter erbjuds att lämna in kompletterande uppgifter. Deadline för inlämning av kompletterande uppgifter är datumet för omtentan. Därefter hänvisas studenter till delkursexaminationen under nästföljande år.

Övriga krav för slutbetyg

För att bli godkänd på kursen krävs att både tentamen och inlämningsuppgifter godkänts. Slutbetyg på kursen baseras på betyget på tentamen.

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.