

Datavarehus – hva er det?

- Bill Inmon (ca. 1992):
 - a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process
- Ralph Kimball (ca. 1996):
 - a copy of transaction data specifically structured for query and analysis
- Data mart:
 - a simple form of a data warehouse that is focused on a single subject
- ETL:
 - Extract, transform and load
- Fact:
 - a value or measurement
- Dimension:
 - reference information that gives context to the facts

Datavarehus CERES (DV)

- Økonomi:
 - Økonomidata (kilde OA) for UiB og UiO overføres daglig til DV. Dekker hovedtyngden av økonomirapportering.
- Studie (STAR):
 - FS-data overføres ukentlig for samtlige FS-institusjoner.
- Forskning:
 - Cristin-data overføres daglig. Hovedvekt på publikasjonspoeng.
- SO: Daglig overføring av opptaksdata fra SO, ukentlig fra NVB
- UiB: Koblede data: Løsning som sammenstiller data på stednivå. Presenterer nøkkeltall, f.eks. Kostnader/publikasjonspoeng.
- UiO: Henter data fra DV til egen løsning i Corporater. Inneholder også data fra de 3 kildene (økonomi/studie/forskning).
- Teknologi i DV:
 - ETL-verktøy: PowerCenter (Informatica)
 - Database: Oracle
 - Rapportverktøy: Tableau

FS datavarehus

- FS datavarehus = en delmengde av FS tilrettelagt for rapportering og analyse
- Ca. 140 FS-tabeller overføres hver lørdag formiddag til DV fra samtlige FS-databaser, dvs. FS01-FS07, UiO, UiB
- Testmiljøet henter fra DMO/test-databaser
- Produksjonsmiljøet henter fra PRD/prod-databaser
- Tilgang til data i DV/Oracle har tidligere vært håndtert ved hjelp av VPD, slik at brukerne bare får tilgang til egen institusjon
- Nå ligger de samme dataene i Tableau (i form av ekstrakter), der tilgangen håndteres med en radkontroll-funksjon
- Tableau-ekstraktene oppdateres hver lørdag formiddag

Forskjeller mellom FS og DV

- FS er et 'studieadministrativt system'
- normalisert datamodell
- designet for å registrere studentinformasjon
- distribuerte databaser
- inputkontroll/constraints for å sikre datakvalitet
- helt ferske data
- komplisert datamodell (> 1000 tabeller og 17.000 kolonner)
- FS-datavarehus er en løsning for analyse/rapportering
- stjernemodeller
- designet for å rapportere/analysere
- en database
- alle data er hentet fra FS (kvaliteten er som i FS)
- ukegamle data (lastes lørdag)
- enkel datamodell (< 100 tabeller og < 2000 kolonner)

Datamodell for gjennomstrømming i FS

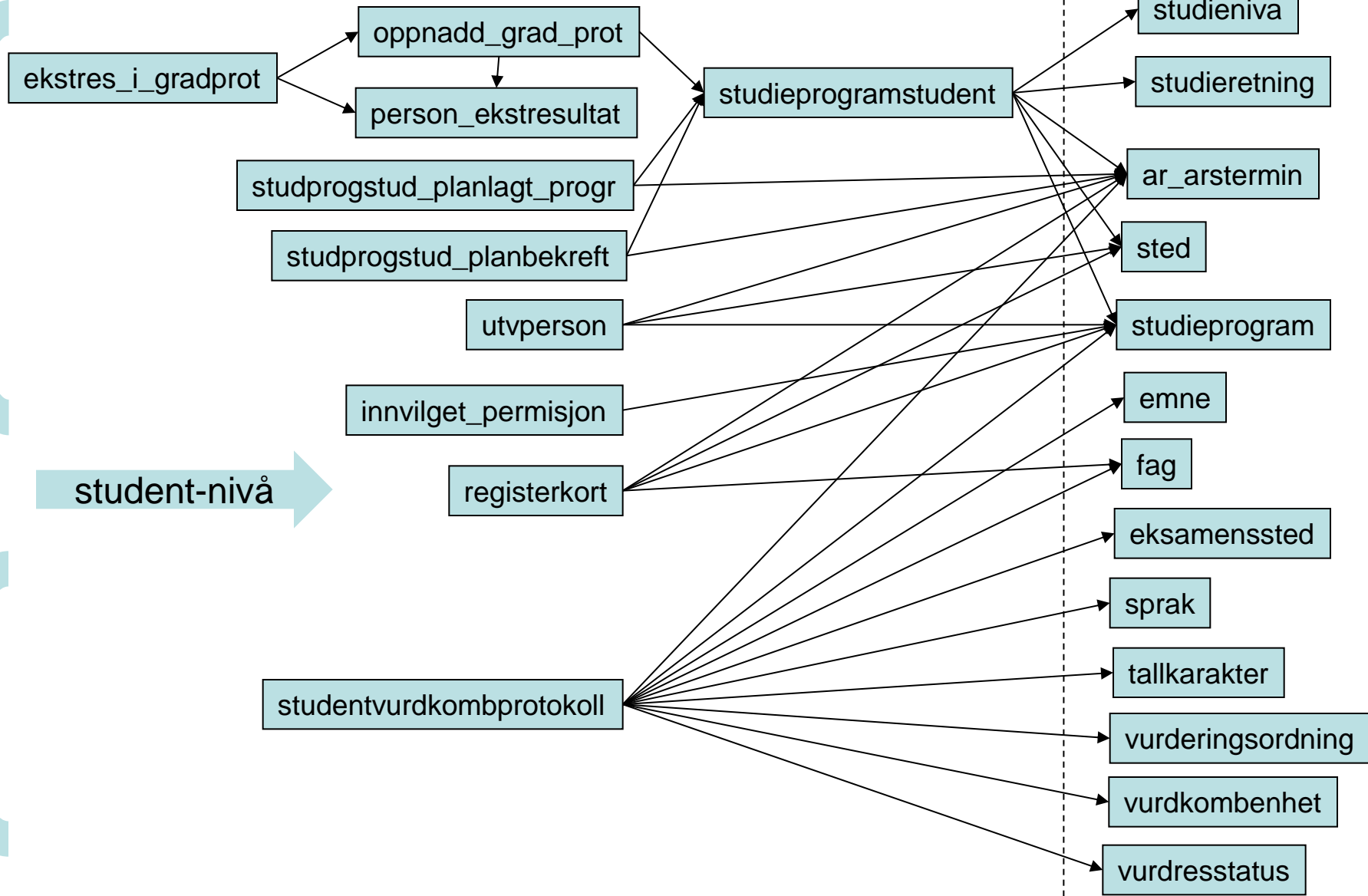
fakta

dimensjoner

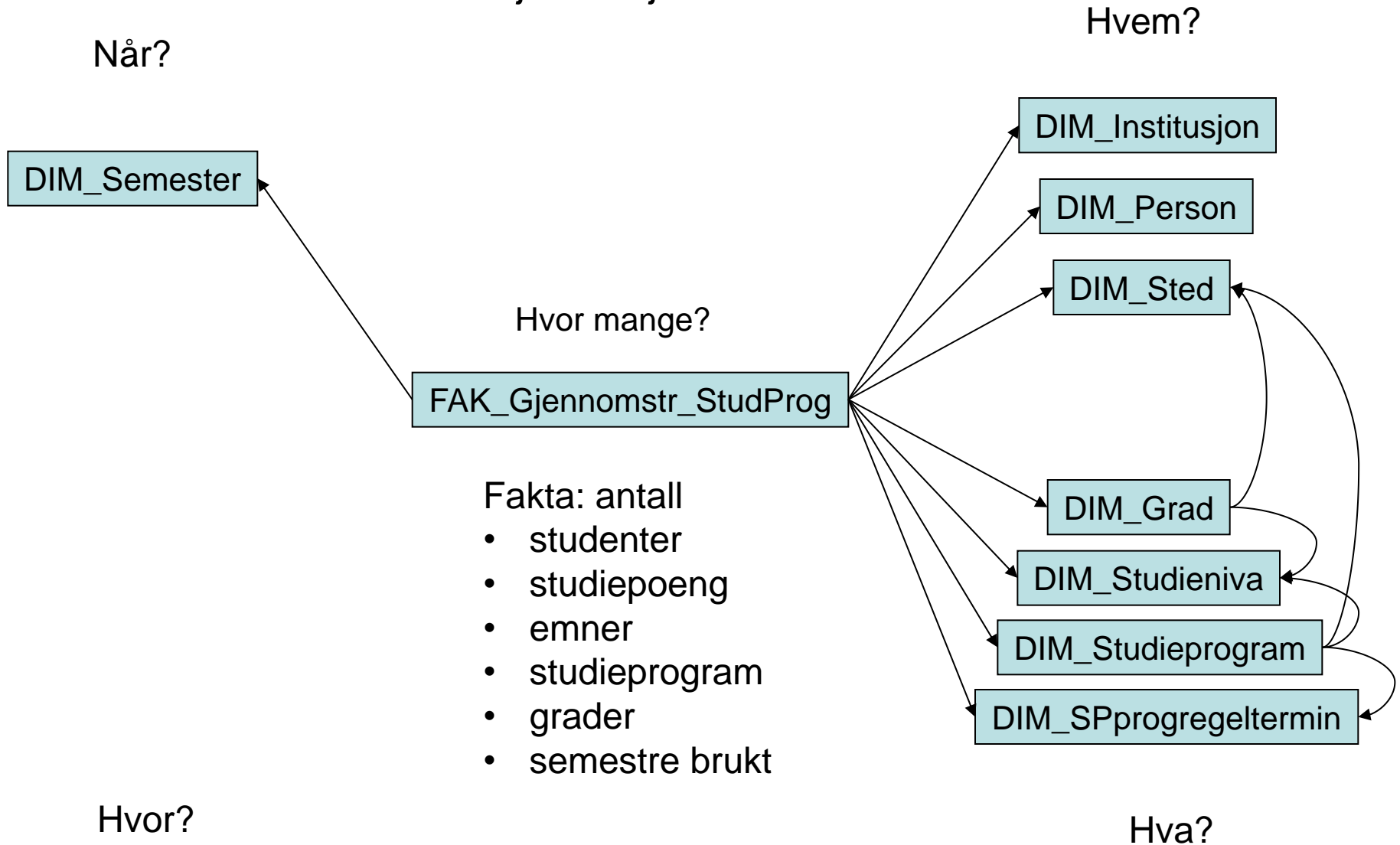
program-nivå

student-nivå

emne-nivå



Datamodell for gjennomstrømming i DV: Stjerneskjema

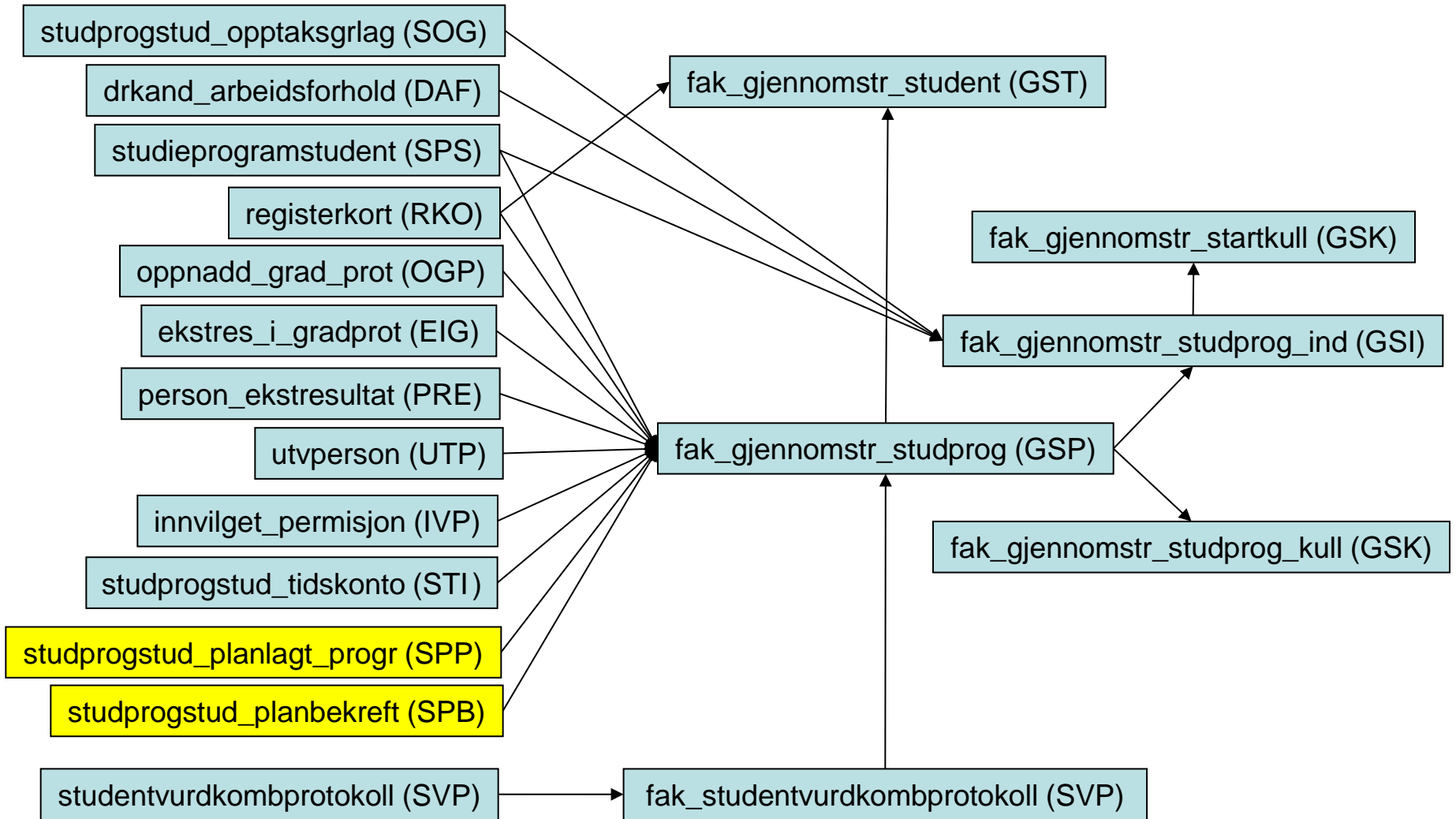


Gjennomstrømming:

Hva skjer på veien fra FS til DV?

- Standard-semester: VÅR/HØST (FS har flere semestre)
- Fra dato til semester (VÅR/HØST) (f.eks. gradopptak)
- For grader/kvalifikasjoner: Januar = Høst, Juli/August = Vår
- Utbredning: Fra-til-semester omdannes til flere rader
- Sammenslåing av tabeller på samme nivå (emne/program/student)
- Aggregering fra emne til program til student
- Prioritert program: Grad-/utdplan- eller primærprogram
- FYLT-felter gir sporbarhet tilbake til kildetabell
- Noen tabeller er primære (=> insert), andre sekundære (=> update)
- Helt nye elementer:
 - Semester-tellere: Kalender, brutto og netto (tar hensyn til permisjon/progresjon)
 - Akkumulerte studiepoeng
 - Faktorer for permisjon, utveksling, progresjon

Dataflyt for gjennomstrømming



Primærkilde (insert)

Sekundærkilde (update)