**Database modeling notes**

**Week: 1 / Topic: Data modeling / PowerPoint: Anniken /Språk: Norsk**

**Data modellering**

**DBMS**

Et database håndteringssytem (DBMS) er et datastyrt system som lar brukere lage og opprettholde en database. DMBS er et programvaresystem (software) som legger til rett for:

* Definering = spesifisere datatyper, strukturer og begrensninger til dataen
* Konstruksjon = lagre dataen på et medium som kontrolleres av DBMS
* Manipulasjon = henting av spesifikk data, oppdatering pga. endring i miniverden, osv.
* Deling = lar flere brukere og programmer få tilgang til databasen samtidig

**Databasesystemet**

* Programmer får tilgang til databasen ved å sende forespørsler om data til DBMS.
* Ved en transaksjon kan noe data bli lest eller skrevet inn i databasen.
* Andre viktige funksjoner ved DBMS inkluderer beskyttelse av database og vedlikehold/bruk over lengre tid (oppnås vha tuning).
* Beskyttelse innebærer systembeskyttelse fra funksjonsfeil og sikkerhet mot uautorisert eller skadelig tilgang.
* DBMS og databasen kombinert kalles databasesystemet.
* Databasesystemet inneholder en DBMS katalog som lagrer beskrivelsen av databasen (eks: datastruktur, typer og begrensninger). Denne beskrivelsen kalles metadata.

**Deling av data og flerbrukets transaksjonsbehandling**

* En flerbruker DBMS må la flere brukere få tilgang til databasen samtidig.
* DBMS må ha programvare fro samtidighetskontroll, slik at når flere brukere prøver å oppdatere den samme dataen, så blir dette gjort på en kontrollert og effektiv måte.
* For eksempel ved salg av flyseter må DBMS sikre at kun en bruker får reservere ett bestemt sete. Dette kalles OLTP (Online Transaksjon Prosessering) og lar flere hundre samtidige transaksjoner utføres per sekund.
* Transaksjoner er utførende programmer som inkluderer akses til en eller flere databaser, for eksempel lesing eller oppdatering av databaseeposter.
* Samtidighetskontroll garanterer at hver transaksjon blir riktig utført eller avbrutt.
* DBMS må sikre flere egenskaper ved transaksjoner, for eksempel isolasjon (hver transaksjon ser ut til å utføres isolert fra andre som utføres samtidig) og atomisitet (alle operasjonene i en transaksjon blir utført eller ingen).

**I Store Organisasjoner vil mange personer være involvert i designet og opprettholdelsen av større database**

* Database administratorer (DBA) – styrer databasen, DBMS og relatert programvare som blir brukt av mange personer.
* DBA autoriserer aksess til databasen, koordinerer bruken, henter program- og maskinvare som trengs og sørger for sikkerhet og god responstid.
* Database designere – identifiserer dataen som skal lagres i databasen og velger passende strukturer for å representere og lagre dataen.
* Sluttbrukere – brukerne som er grunnen til at databasen eksisterer.
* System analytikere og applikasjonsprogrammerere – utvikler programvaren og bestemmer kravene til brukerne, tester, debugger, osv.
* Andre personer er ansvarlig for design, utvikling og drift av DBMS programvaren og systemmiljøet. Disse er som regel ikke interessert i selve databaseinneholdet: DBMS systemdesignere, verktøyutviklere, personell for drift og vedlikehold.

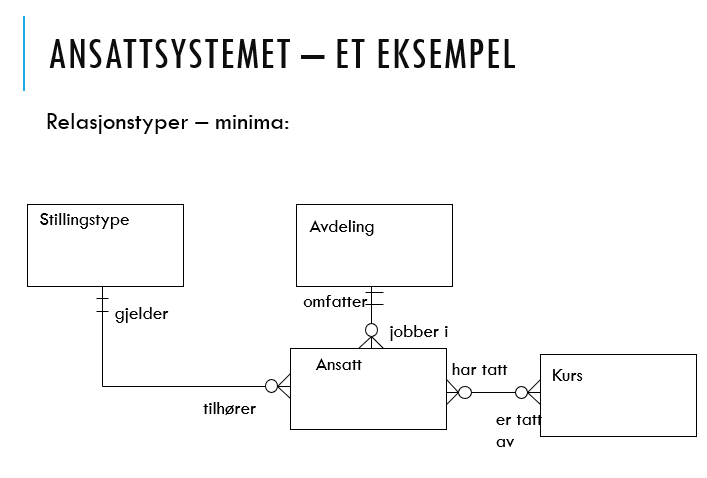
**Fordeler ved å bruke en DBMS**

* Kontroll av redundans
* Hver bruker opprettholde sine egne filer, noe som fører til duplikater av samme fil.
* Gir bortkastet lagringsplass og mer arbeid for å oppdatere dataen.
* DBMS bruker redundansskontroll for å sikre at disse duplikatene er riktige.
* Begrenser uautorisert tilgang
* Gir vedvarende lagring for programobjekter – et kompleks objekt skrevet i C++/Java kan lagres permanent i en objektorientert DBMS, slik at den «overlever» terminering av programmet og kan senere direkte hentes av et annet program.
* Gir lagringsstruktur og søketeknikker for effektiv forespørsel.
* Gir backup og gjenoppretting – DBMS legger til rette for gjenoppretting etter feil i maskinvare eller programvare.
* Gir grensesnitt for flere brukere.

**Entitetstype**

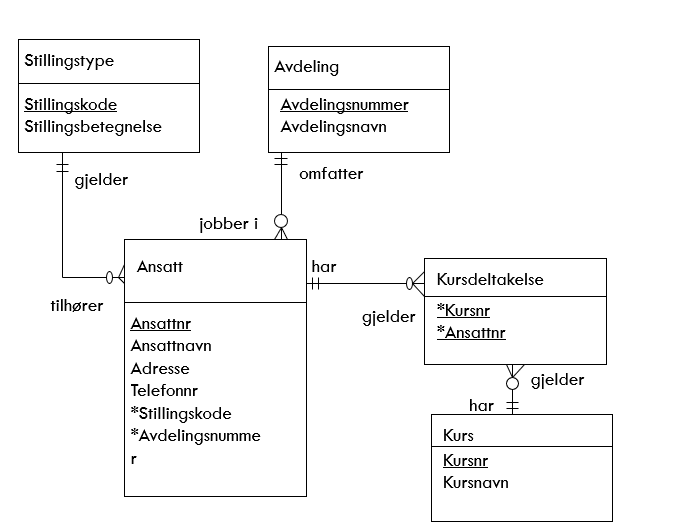
**Entitet**: Et eller annet innenfor vårt interesseområde som vi ønsker informasjon om.

**Entitetstype**: En samling entiteter der alle entitetene oppfyller definisjonskriteriene for entitetstypen.



**Hoved- og fremmednøkler**

* For at datamodellen skal kunne brukes som grunnlag for en database må basen ha en form for «kobling» mellom de ulike entitetstypene.
* Vi må bringe inn hovednøkkel og fremmednøkkel-begrepet.
* Hovednøkkelen – unik.



**Identifikator = Primærnøkkel = Hovednøkkel**

* Vanligvis består identifikatoren av ett eneste attributt
* De forekommer identifikatorer som er sammensatt av to eller flere attributter = sammensatt identifikator

**Identifikator – Definisjon**

En indentifikator for en enititetstype er et attributt eller en gruppe av attributter som velges ut til å identifisere de ulike forekomster av en entitetstype, og som oppfyller følgende egenskaper: Entydighet, irredusibilitet og entitetsintegritet.

**Entydighet** = Alle verdir for attributtet er forskjellige

**Irredusibilitet**: Aktuelt ved sammensatt identifikator: Skal ikke kunne kutte ut ett eller flere attributter og likevel kunne bruke de gjenværende som identifikator

**Entitetsintegritet**: Identifikatoren må ha en verdi.

**Week: 2 / Topic: Module 2 – Database Basics / PowerPoint /Språk: Engelsk**

**Data modeling and Database Application**

**Database is…**

* From 1960s
* Is an efficient technology to manage data
* Is an important branch in computer science subject
* The infrastructure of information systems (digital systems)
* Helps computer applications in different fields.

**4 Concepts in Database**

* Data
* Database
* Database Management System (DBMS)
* Database System

**Data is…**

* The basic component stored in the database.
* Symbol record to describe objects.
* Abstract & communication
* Types of data
* Number, string, text, image, audio, video, profiles, etc.
* Meaning of the data is semantics. In CS, data has to combine with its semantic.

**Database is…**

* An organized, sharable collection of large amounts of data, generally stored in the computer for longer period.
* Characteristics:
* Data are organized, described and stored according to some data models.
* Can be shared among different types of users.
* Redundancy is low.
* Independency is high.
* Easy to expand (multiple dimensions, keep this in mind).

**Database Management System (DBMS) is…**

* Data management software between user and OS
* A fundamental software, large, complex software system.
* Used for:
* Organize and store data in an efficient way.
* Easy to access and maintain the data.
* Functions part 1:
* Data definition

1. Data Definition/Description Language (DDL)
2. Define data objects in database

* Data organization, storage and management

1. Classify different data
2. Confirm data structure and access method
3. Implement data connections/relationships
4. Provide multiple access methods to improve efficiency

* Functions part 2:
* Data manipulation

1. Data Manipulation Language (DML)
2. Implement basic database operations (CRUD: Create, Reade, Update, Delete)

* Transaction and operation management in database

1. Unified management and control for database create, operate and maintain by DBMS
2. Support data security, integrity, concurrent use from multiple users.
3. System recovery after failure

* Functions part 3:
* Database maintenance

1. Initial installation and migration
2. Reconfiguration
3. QoS monitoring, analysis, etc.

* Others

1. Communications with other local software
2. Communications with other nodes on the web
3. Data transformation with other DBMS
4. Interoperability between heterogeneous systems

DB

Application

Application development tool

OS

DBMS

Administrator

User

User

User

…

**Database System**

* Contains
* Database
* DBMS & application development tools
* Application
* Administrator
* Structure

**History**

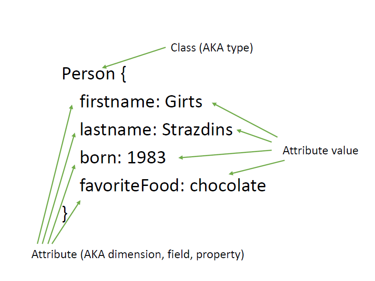
* Database management technology is
* To classify, organize, code, store, access and maintain data
* Center of data processing
* Phases
* Manual Management
* File system
* Database
* Motivations & Constraints
* Application
* Computer hardware
* Computer software

**Characteristics of DB**

* Data is structured
* Systematical structured
* High sharing, low redundancy, easy to expand
* High indenpendency
* Complete physical independency
* Relative logical independency
* Unified management and control by DBMS
* Control: Security, integrity, concurrency, recovery

**Week: 3 / Topic: Module 2 – Data Modeling / PowerPoint /Språk: Engelsk**

**Data Modeling and Database Application**



**Data models**

* Are abstract descriptions of data features from objective world.
* Should:
* Describe the world in a relatively real way
* Easy to be understood
* Easy to be implemented on computers
* Are the center and foundation of database system
* Elements in data models
* Basic data models
* Hierarchical models
* Network models
* Relational models

**Elements in data models**

* Data structure
* Data manipulation
* Constraints for integrity

**Data structure of data models**

* To describe objects of database and the relationship between objects

**Content**

* Related to the object type, value, feature.
* Related to the relationship between data.

Data structure is a description of system static features

**Data Manipulation**

* The operation set that can be executed for all object instances (both types and values)

1. Operation types
2. Operation rules

**Operation types**

* Access (Query)
* Update (Add, delete, modify)

**Definition of data manipulations in data models**

* The exact meaning of operation
* Operation rules
* Manipulation Language

Data manipulation is the description for system dynamic features

**Constraints for data integrity**

* A collection of integrity rules
* Integrity rules: constraints and dependency rules for data and relationships in data models.
* Used to limit the database state and the state transfer according to data models in order to guarantee the correctness, effectiveness and compatibility of data.
* Basic integrity constraints
* Advanced constraints to reflect concrete application requirement

**Basic concepts need to be expressed in data models**

* **Entity**

1. Things exist in objective world and can be identified
2. Can be people, objects, events or abstract concept

* **Attribute**

1. Features that possessed by entities
2. One entity can be described by several attributes

* **Key**

1. The attribute(s) which can Uniquely identify entity

* **Domain**

1. Value range

* **Entity type**

1. An entity type is a template for entity type instances, including entity name and attribute sets

* **Entity set**

1. The collection of entites with the same entity type (Value)

* **Relationship**

**Data Models**

Elements in data models

Basic logic data models

* Hierarchical models
* Network models
* Relational models

Hierarchical model

* Earliest data model in database
* Typical DBMS is information Management System (IMS) from IBM
* Uses tree structure to describe different entity types and relationships between entities

Two requirements

* Exist only one node has no parent node; this is called root node
* All the nodes except root node exist only one parent node

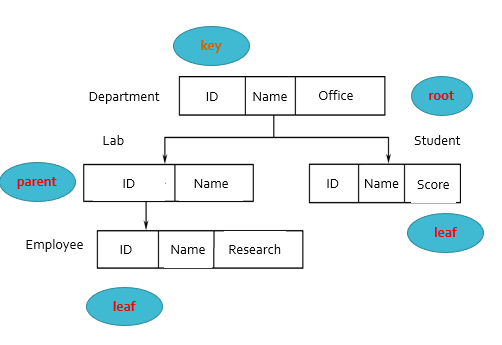
Terminology

* Root node, parent node, sibling node, leaf node.

Basic Data Models – Hierarchical Model

Characteristics of Hoerarchical Models:

* Parent node for a given node is unique
* Only deal with the 1 to N relationship between entities
* Every record can define sorting attribute, it is called key attribute.
* All the recordings can show their complete semantics only based on the query the whole path
* There is no child value can independently exist without parent recording



Data manipulation for hierarchical models

* Create, Read, Update, Delete (CRUD)

Integrity Constraints

* No create operation can apply to child node values without parent node values
* Delete parent node values mean all the child nodes values are deleted at the same time
* Update operation should apply all the related values to keep data consistency

Advantages

* This model is simple and clear
* High efficiency for query, higher than relational model, similar to network model
* Good integrity support

Disadvantages

* Difficult to describe N to N relationships
* More constraints for create and delete operations. Application development is complex
* To read the values of child nodes have to go through parent node

Storage structure

* Store values, and structure together
* Adjacency method, Linked method

DMBS Product

* IBM IMS

**Basic Data models – Network Model**

Network database system takes network model as data organization method.

DBTG system (Data base Task Group)

* Also called CODASYL system (Conference on Data System Languages)
* A system solution proposed by DBTG in 1970s

Network model has two requirements:

* It is allowed for more than one node to have no parent node
* One node can have more than one parent node

Description:

* Entity type
* Attributes
* Relationships

Differences between network models and hierarchical models:

* Network model allows multiple nodes without parent node (multiple root nodes).
* Network models allow nodes to have multiple parent nodes.
* Network models allow multiple (combinational) relationships between two nodes. (two persons can be both work and family relationships).
* Network models can describe the real world more directly.
* Hierarchical model is a special case of network models.

Advantages:

* Can describe the real world in a more flexible way, because one node can have several parent node.
* Good efficiency for query.

Disadvantages:

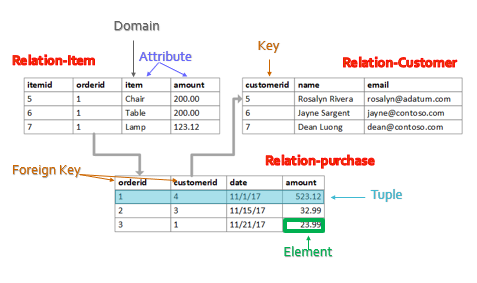
* The structure is more complex, especially, with the application environment expanding. It is not easy for users to handle. (Tax system?)
* DDL, DML are complex, difficult to use by users.
* The relationship is realized by path, we have to know the concrete data structure.

**Basic Data Models – Relational Model**

* Relational database uses relational model as the data organization method.
* From the perspective of users, the logical data structure of relational model is a two-dimensional table, including rows and columns.

Et bilde som inneholder bord

Automatisk generert beskrivelse



* Data manipulation is a set of collective operations. Operation object and results are all relations.
* CRUD
* Storage path is hidden from the users.
* They just need to tell “**what to do**”, no explanation on “**how to do**”.

Advantages

* Based on mathematical theories.
  + Simple concepts:
    - Entities and all types of relationships can all be described as relations.
    - Query results for data are relations as well.
    - Hidden path
      * High data independency, better security.
      * Simplified the workload for programmers and database development.

Disadvantages:

* Low efficiency for query compared with structured data models.
* Have to optimize user query requirements in order to improve efficiency. More difficult to develop DBMS

**Week: 4 / Topic: Module 3 SQL / PowerPoint /Språk: Engelsk**

**Week: 5 / Topic: Module 4 Database security / PowerPoint /Språk: Engelsk**