Fachhochschule Vorarlberg
University of Applied Sciences

Software Architecture Mining in Relational DBMSs

Bernd G. Wenzel

Overview

- Some Observations
- From DD to Logical DB Structure
- From LDS to Model
- Limitations of this Approach
- Conclusions
- Questions and Answers

Some Observations (1)

- Model-based techniques are widely accepted for software design
  - Acceptance still growing
- Model-based techniques are getting accepted for software maintenance
  - Provided they have been used for designing the software
- Model-based techniques are considered one-way-streets from models to implementation
- Software systems consist technology artefacts and of representations of domain elements
  - Hard distinction between the two
- Relational Databases are everywhere
  - and with them their DD (Data Dictionary)

Some Observations (2)

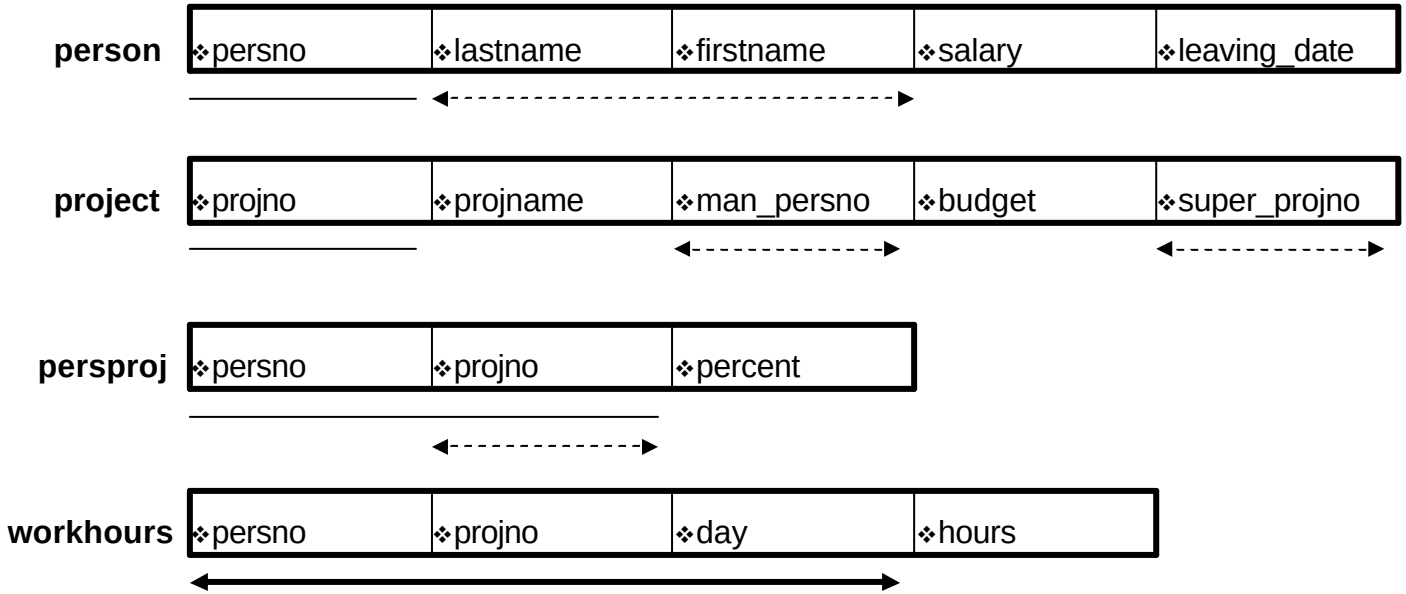
- Persistent data/objects are most likely representations of domain elements
  - and not technology artefacts
- On the logical level relational databases have very few concepts
  - Making it simple to filter out the artefacts of relational technology

From DD to Logical DB Structure (1)

- The content of a relational data dictionary
  - Mandatory
    - Tables
    - Columns of tables
      - » With their datatypes
    - Indexes
      - » Unique
      - » Non-unique
  - Optional
    - Primary key definitions
    - Uniqueness constraints
    - Referential integrity constraints
    - NULL value constraints
    - General constraints

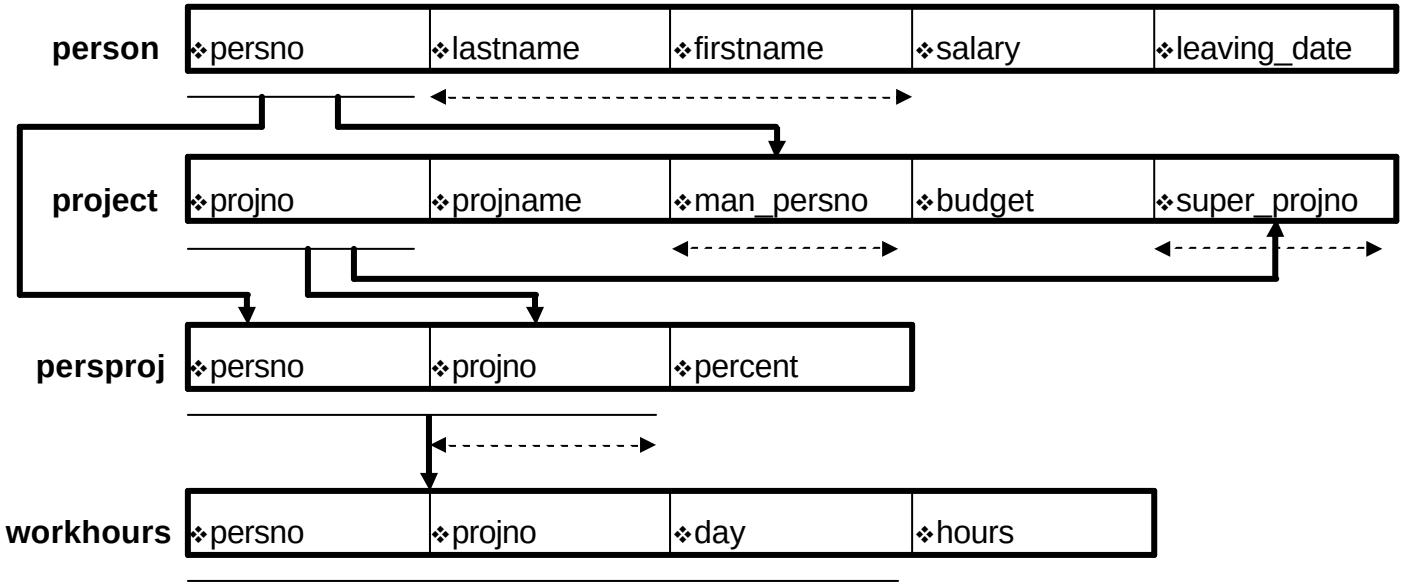
From DD to Logical DB Structure (2)

- Process Example
  - Completing uniqueness information
    - manually



From DD to Logical DB Structure (3)

- Process Example (cont.)
  - Identification of references
    - Supported by automated search

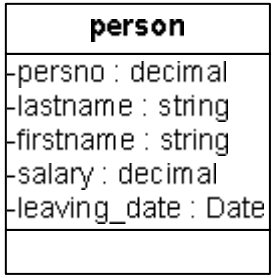


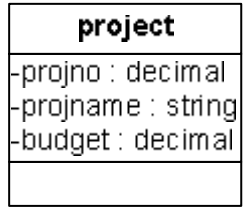
From DD to Logical DB Structure (4)

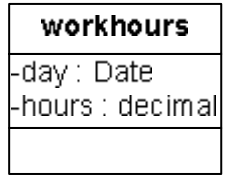
- Observation
  - Databases in operation aren't perfect
- Consequence
  - Use all information sources, such as
    - This process
    - Domains and data types
    - Inactive constraints
    - etc.
- Quality of result
  - Depends on naming conventions applied
    - Adaptations possible
  - Cannot be guaranteed
    - Human supervision needed

From LDS to Model (1)

- Process Example
  - Every table, the primary key of which does not only consist of foreign keys, becomes a class
    - Foreign key attributes are removed

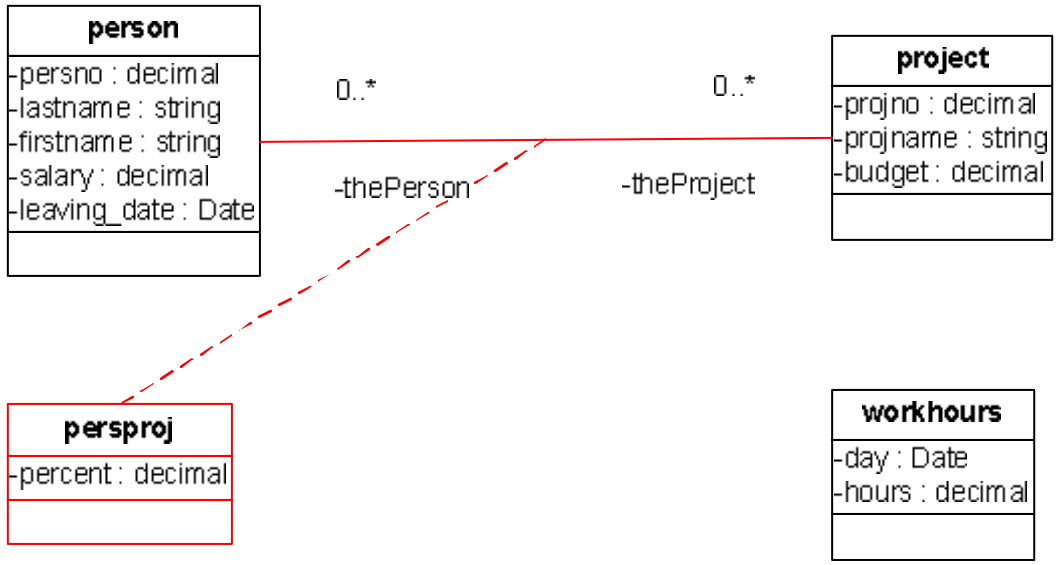






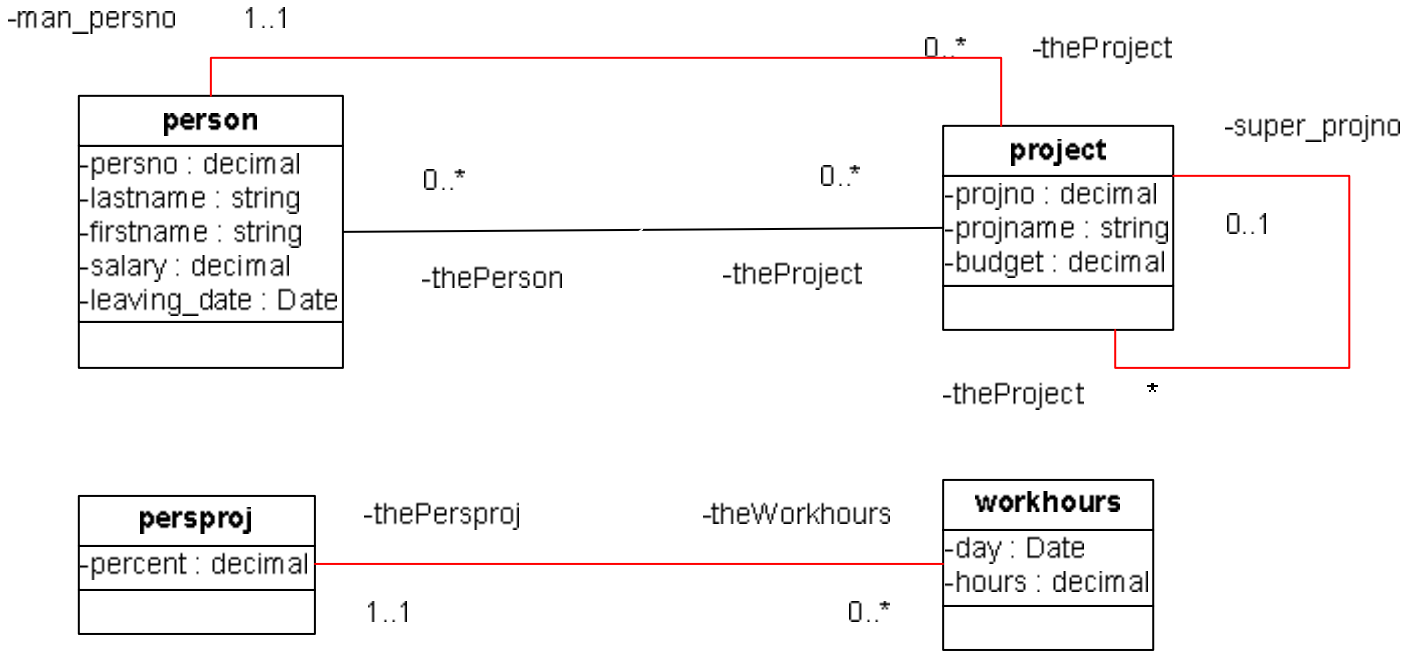
From LDS to Model (2)

- Process Example (cont.)
  - Every table, the primary key of which consists only of foreign keys becomes an association or an association class (if it has additional attributes or a foreign key reference to ist primary key)



From LDS to Model (3)

- Process Example (cont.)
  - The remaining foreign key references become associations



From LDS to Model (4)

- Process Example (cont.)
  - Tables with identical keys are candidates for generalization/specialization relationships
    - No picture
- Admitted facts and observations
  - Generated model is incomplete
    - E. g. regarding public or private attributes
  - Generated model may be wrong
    - Quality depends on quality of database
      - » GIGO Principle
  - Generated model contains the most important classes and associations usually
    - and not much noise

Limitations of this Approach

- No guarantee of success
  - No push button solution
  - But potential for significant computer support
- Data modelling only
  - No complete class model
  - No behaviour
- But a potential condensation point for additional information
  - Use the class model we get
  - Apply a sanity check and do the required changes
  - Search for the places where the original constructs were used
  - Adapt and connect these code fragments

Conclusions

- We have seen an approach to filter information relevant for system modernisation from the DD's of relational databases
  - It doesn't work under all circumstances
    - Needs human supervision
  - It can be adapted to other database technologies
  - It is restricted to mining components of software architectures
    - No support for complete systems
  - Its basics are known and have been used successfully since about 15 years

Questions and Answers

The most stupid questions are those never asked

There are no stupid questions, only stupid answers