

Zusammenführung der konzeptuellen Modelle für operative und analytische Daten in einem logischen Modell, dargestellt am Beispiel des Precision Dairy Farming

Christian Schulze¹, Joachim Spilke¹, Wolfgang Lehner²

¹ Arbeitsgruppe Biometrie und Agrarinformatik
Martin-Luther-Universität Halle-Wittenberg
Ludwig-Wucherer-Str. 82-85
06108 Halle

² Lehrstuhl für Datenbanken
Institut für Systemarchitektur
Technische Universität Dresden
01062 Dresden
christian.schulze@landw.uni-halle.de
joachim.spilke@landw.uni-halle.de
lehner@inf.tu-dresden.de

Abstract: Die als Ergebnis der konzeptuellen Datenmodellierung im Rahmen des Precision Dairy Farming vorliegenden getrennten Modelle für operative und analytische Daten sind in ein logisches Modell zu überführen. Wir propagieren dabei den Entwurf eines gemeinsamen relationalen Modells für beide Datensichten. Nur so kann Redundanz und Inkonsistenz auf Modell- und Datenebene vermieden werden. Anhand eines Beispiels wird die Vorgehensweise erläutert.

1 Motivation

Grundlage der Konzepte des Precision Agriculture ist ein angepasstes Informationsmanagement, bei dessen praktischer Umsetzung der Datenmodellierung eine Schlüsselrolle zukommt. Hierbei liegt ein besonderes Interesse in der Unterstützung einerseits kurzfristiger (operativer) und andererseits strategischer (analytischer) Entscheidungsprozesse durch konsistente Datenspeicherung und einer zeitkritischen Datenbereitstellung. Beide Nutzungsarten münden in getrennten Datensichten, die entsprechend im Modellierungsprozess zu berücksichtigen sind. An einem Beispiel des Precision Dairy Farming (PDF) wird im vorliegenden Beitrag der logische Datenbankentwurf sowohl für die operative als auch für die analytische Datensicht ausgehend von entsprechend getrennten konzeptuellen Datenmodellen [SSL05] hin zu einem gemeinsamen relationalen Modell gezeigt. Die Allgemeingültigkeit der getroffenen Aussagen sichert die Anwendbarkeit über das PDF und Precision Agriculture hinaus.

2 Konzeptuelle Modellierung

Im Ergebnis der konzeptuellen Datenmodellierung im Rahmen des PDF liegen zwei getrennte Datenmodelle, zum einen für eine operative und zum anderen für eine analytische Datensichtweise vor [SSL05]. Der operativen Datenmodellierung kommt die Aufgabe der Abbildung atomarer Daten zu, während im Rahmen der analyseorientierten Datenmodellierung neben atomaren Daten und Datenaggregationen auch Analysestrukturen im Sinne möglicher Wege der Auswertung für den Anwender abgebildet werden müssen. Entsprechend nutzen wir für die Modellierung operativer Daten das Entity-Relationship-Modell (E/RM) [EN03] und für die Abbildung einer analyseorientierter Sicht das multidimensionale Entity-Relationship-Modell (mE/RM) [Sa98] mit von uns als notwendig erachteten Erweiterungen [SSL05]. Abbildung 1 stellt beispielhaft Krankheitsbehandlungen an Kühen bzw. Eutervierteln als E/RM dar. Die Behandlung ist über Beziehungstypen zwischen den Entitätstypen „Zeitpunkt“, „Person“, „Krankheitsdiagnose“ und „Euterviertel“ bzw. „Kuh“ modelliert. Zur Vereinfachung wurden weitere Sachverhalte (bspw. Medikamentengaben) vernachlässigt. Dieses E/RM bildet den detaillierten Sachkontext für operative Aufgaben ab. Abbildung 2 zeigt eine analyseorientierte Sicht auf den gleichen Kontext, modelliert mit dem von uns erweiterten mE/RM. Statt der detaillierten Unterscheidung in Tier- und Euterviertelbehandlung ist in der analyseorientierten Sicht die konsolidierte Betrachtung von Behandlungen auf ein Tier inklusive Euterviertelbehandlungen von Interesse.

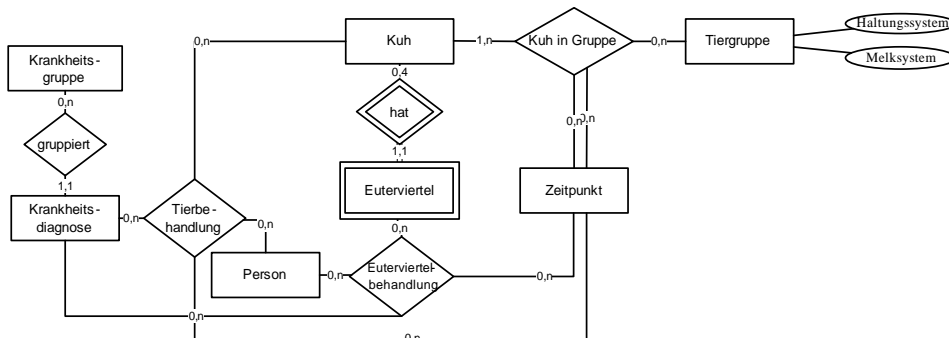
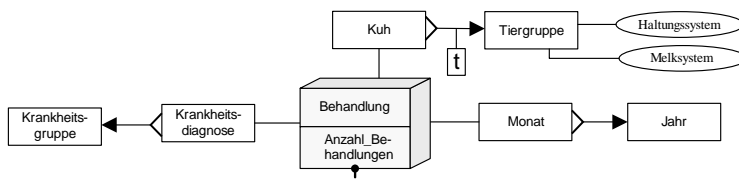


Abbildung 1: E/RM für Tierbehandlungen



Aggregationstyp flow
Metrik: SELECT kuh_nr, diagnose, monat, sum(anzahl_behandlungen) AS anzahl_behandlungen FROM (SELECT tierbehandling.kuh_nr AS kuh_nr, tierbehandling.diagnose AS diagnose, (date_part('year', tierbehandling.zeitpunkt)) || '-' || date_part('month', tierbehandling.zeitpunkt) AS monat, count(tierbehandling.*) AS anzahl_behandlungen FROM tierbehandling GROUP BY tierbehandling.kuh_nr, tierbehandling.diagnose, monat UNION SELECT euterviertelbehandling.kuh_nr AS kuh_nr, euterviertelbehandling.diagnose AS diagnose, (date_part('year', euterviertelbehandling.zeitpunkt)) || '-' || date_part('month', euterviertelbehandling.zeitpunkt) AS monat, count(euterviertelbehandling.*) AS anzahl_behandlungen FROM euterviertelbehandling GROUP BY euterviertelbehandling.kuh_nr, euterviertelbehandling.diagnose, monat) AS behandlung_union GROUP BY kuh_nr, diagnose, monat;
Beschreibung Anzahl der Krankheitsbehandlungen je Tier und Monat
Maßeinheit: Monat¹

Abbildung 2: Erweitertes mE/RM für Tierbehandlungen

Analysegegenstand ist die Anzahl der Behandlungen je Kuh, Monat und Krankheit, modelliert als Kennzahl der Faktbeziehung Tierbehandlungen. Die eindeutige Definition der Kennzahl erfolgt über ein entsprechendes Eigenschaftsfenster. Mögliche Auswertungspfade sind über Klassifikationsstufen und ihre Beziehungen mit entsprechender Kennzeichnung temporaler Restriktionen dargestellt. Damit bildet das erweiterte mE/RM den Sachkontext in einer analyseorientierten Sichtweise ab.

3 Relationale Modellierung

Zweiter Schritt im Datenbankentwurfsprozess ist die Transformation der konzeptuellen Datenmodelle in ein logisches, in unserem Fall relationales Modell [EN03]. Bisherige Anwendungen gehen von einer strikten Trennung der operativen und analytischen Ansätze aus. Bei Betrachtung beider konzeptueller Modelle fällt jedoch eine Vielzahl struktureller Gemeinsamkeiten auf. Wir propagieren deshalb eine Zusammenführung beider Datensichten in einem gemeinsamen relationalen Datenmodell. Die Transformation konzeptueller Notationselemente, sowohl des E/RM als auch des mE/RM, in das relationale Modell erfolgt anhand weitestgehend standardisierter Transformationsregeln [EN03, GB01]. Für die relationale Umsetzung der von uns neu eingeführten Notationselemente sei aus Platzgründen auf [SSL05b] verwiesen. Ergebnis der gemeinsamen relationalen Umsetzung beider konzeptueller Modelle ist ein relationales Modell mit detaillierter Grundstruktur, entsprechend den Anforderungen des operativen Modells und zusätzlichen relationalen Konstrukten zur Abbildung der analyseorientierten Sichtweise (Abb. 3). Die Entitäts- und Beziehungstypen des operativen Modells werden in entsprechende Relationen überführt. Darauf aufbauend sind für eine adäquate Abbildung der analytischen Sicht fehlende relationale Konstrukte zu definieren, so dass Redundanz und Inkonsistenz durch beiderseitige Nutzung identischer Modellteile vermieden wird. Das Attribut „Zeitpunkt“ wird als Datentyp „timestamp“ (Datum + Zeit) modelliert, so dass die Klassifikationsstufen „Monat“ und „Jahr“ lediglich als Sicht auf die Relation „Zeitpunkt“ angelegt werden müssen. Um den bei der Transformation semantikreicher multidimensionaler Notationselemente in das semantikarme Relationenmodell zwangsläufig auftretenden Informationsverlust zu vermeiden, muss parallel zum Datenmodell ein Metamodell implementiert werden. Nur so können Meta-Informationen zu Analysegegenstand und Analysestrukturen bewahrt werden.

4 Zusammenfassung und Ausblick

Die dem PDF inhärente duale Datennutzung für operative und analytische Aufgaben hat entsprechend in der Datenmodellierung Beachtung zu finden. Die damit im Ergebnis des konzeptuellen Datenbankentwurfs vorliegenden getrennten Datenmodelle sollten jedoch im logischen Entwurfsschritt zu einem gemeinsamen Modell zusammen geführt werden. Nur so ist Redundanz und Inkonsistenz auf Modell- und damit auch auf Datenebene zu vermeiden. Bei der physischen Implementierung des Modells muss jedoch in Abhängigkeit von Datenbankmanagementsystem, Datenvolumen und Nutzungsverhalten eine Entscheidung über die Art der Umsetzung der zusätzlichen analytischen Modellteile z.B.

in Sichten, materialisierte Sichten oder partitionierte Tabellen erfolgen.

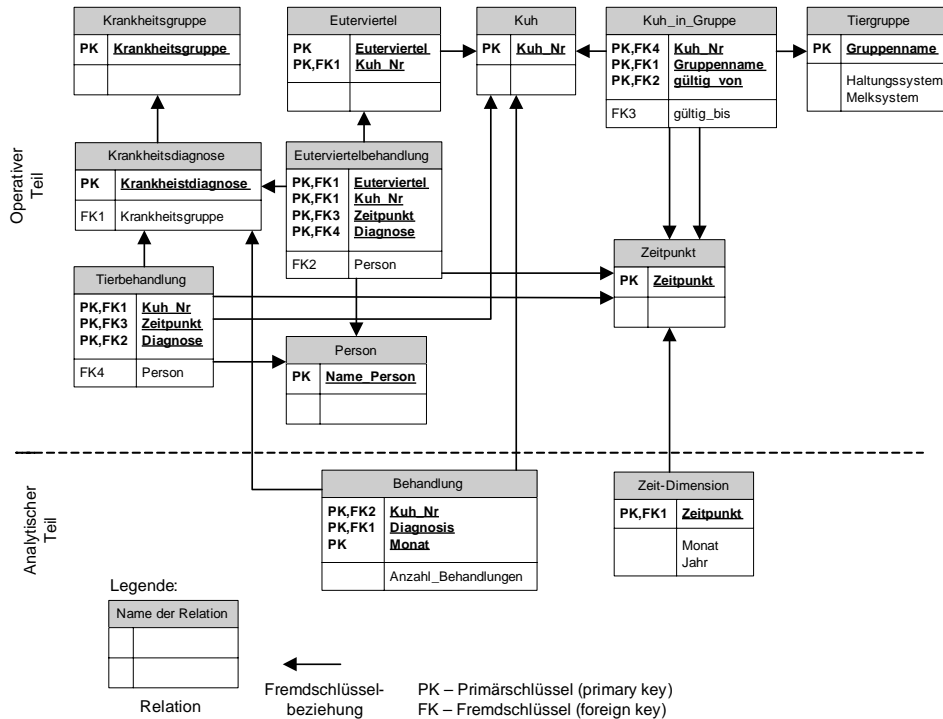


Abbildung 3: Relationales Modell für Tierbehandlungen

Danksagung

Dieses Projekt wird durch die Deutsche Forschungsgemeinschaft (DFG) gefördert.

Literaturverzeichnis

- [EN03] Elmasri, R., Navathe, S.B.: Grundlagen von Datenbanksystemen. 3. Pearson-Studium, München, 2003.
- [GB01] Günzel, H., Bauer, A.: Data-Warehouse-Systeme. dpunkt, Heidelberg, 2001.
- [Sa98] Sapia, C., Blaschka, M., Höfling, G., Dinter, B.: Extending the E/R Model for the multi-dimensional paradigm. In: Proceedings of the 1st Intl. Workshop on Data Warehouse and Data Mining (DWDM'98), LNCS, 1552, Springer, 1998; S. 105-116.
- [SSL05a] Schulze, C.; Spilke, J.; Lehner, W.: Hybred modeling of operational and analytical data at Precision Dairy Farming. In: Tagungsband EFITA/WCCA2005, 25.-27.07.2005, Univ. de Trás-os-Montes e Alto Douro, Vila Real, Portugal, 2005; S. 960-965.
- [SSL05b] Schulze, C.; Spilke, J.; Lehner, W.: Data Modeling for Precision Dairy Farming within the Competitive Field of Operational and Analytical Tasks. Computers and Electronics in Agriculture, 2005 (eingereicht).