

Aufbau und Pflege von Kunden-Ids im Finanzdienstleistungsgewerbe

Thomas Rüdiger

KSFE

Dortmund

01.03.2002

Deduplikation ist die Basis für erfolgreiches CRM

Hintergrund

- Oft um 12 % einer Kundenbasis sind Dubletten
- Dubletten sind Mehrkosten fürs Unternehmen
- Kunden wollen gezielt kontaktiert werden

Probleme

- gelegentlich Kunden-Id nicht vorhanden
- Dubletten lassen sich nicht erkennen (Undermapping)
- Die Dublettensammenführung ist ungerechtfertigt (Overmapping)

Lösung

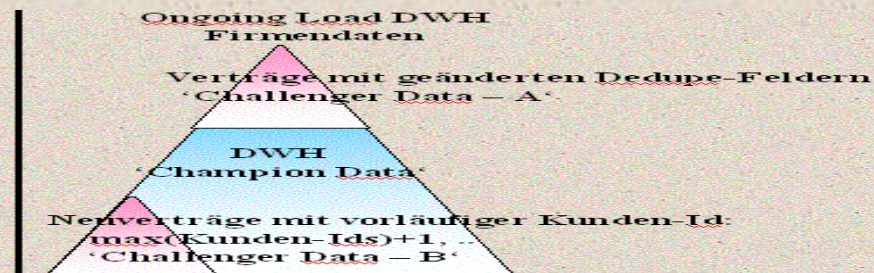
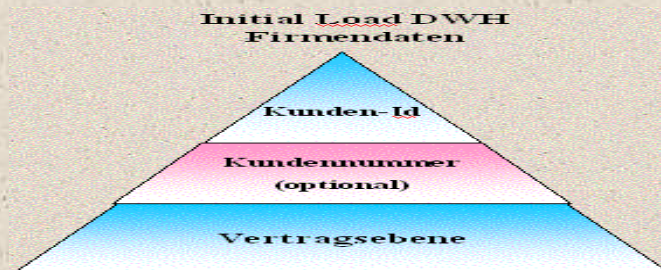
- Datenvorwäsche/Adreßbereinigung
- SAS-Spedis-Algorithmus

Der Kunde hat viele Gesichter, in der Regel eins

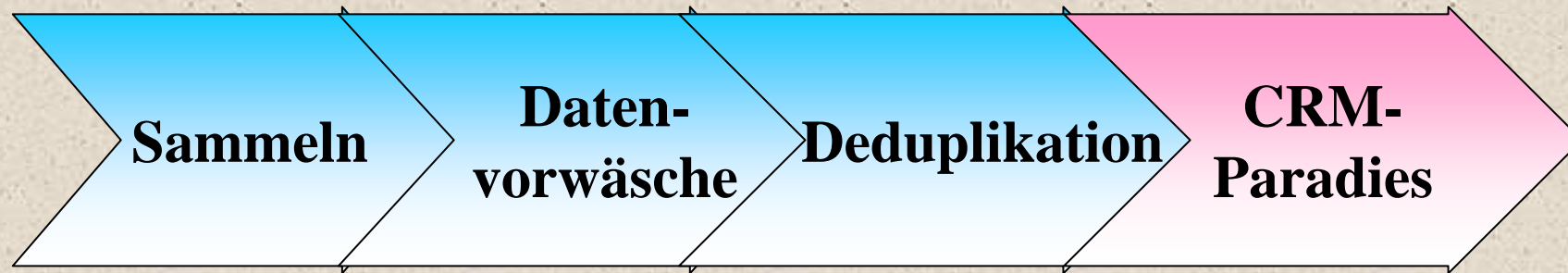
Deduplikation=hierarchische n:1-Abbildung unter Berücksichtigung möglicher textlicher Unschärfen in den untersuchten Feldern



Name	Straße- Hausnummer	PLZ	Ort	Geburtsdatum
Schäfer, Manfred	Blumenstr. 7	50678	Köln- Deutz	01.01.1900
Schäfer, Manni	Blumenweg 7	50678	KOELN	05.01.1965
Schäfer, M.	Blumenstr. 7a	5000	Keoln	05.01.1956



Deduplikation verläuft in Schritten



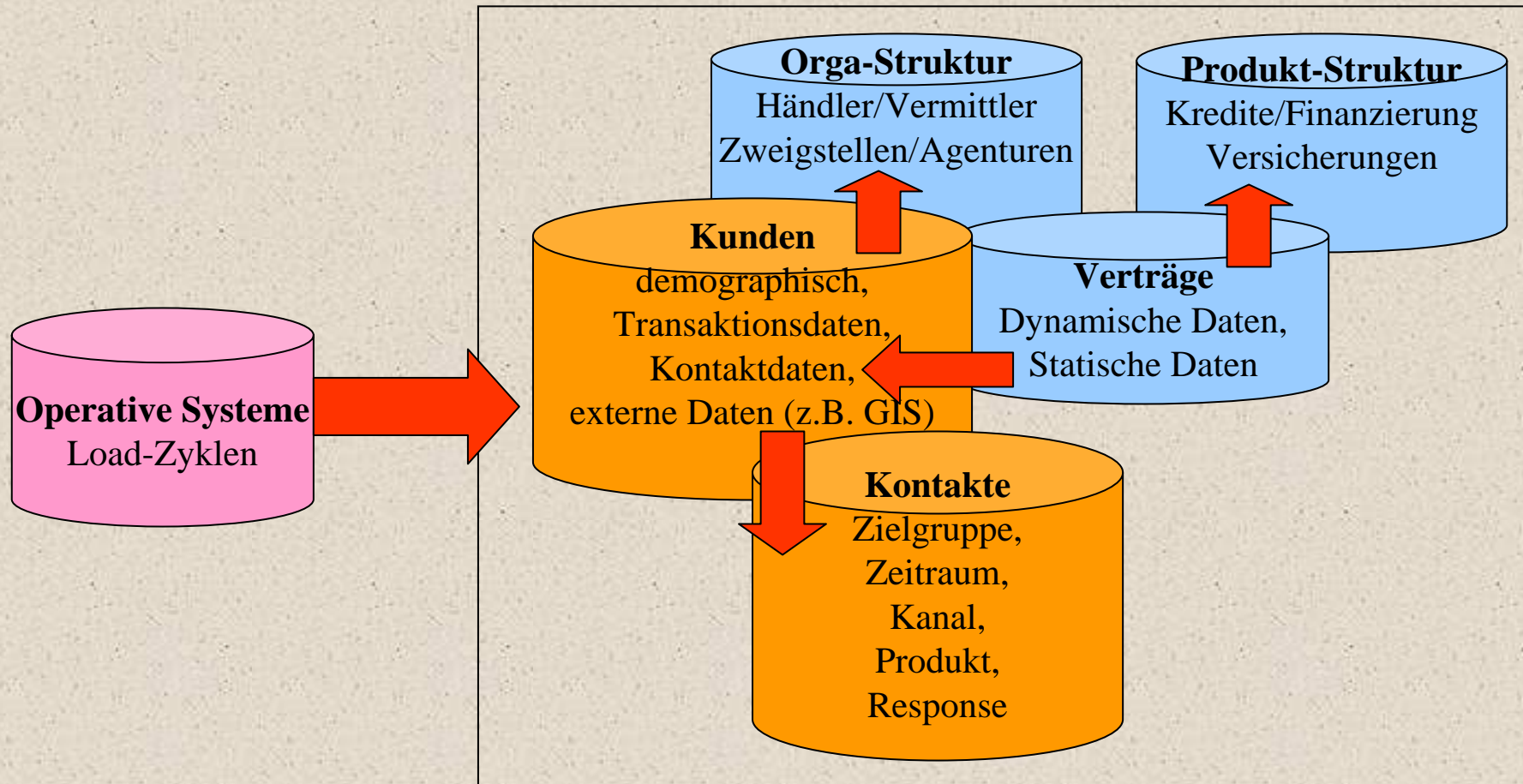
Datenarchitektur
Feldstrukturen
Adreßhistorie

Adreßbereinigung
Datenregeln

Spedis-
Algorithmus
Nachfragen
und bestätigen
Kunden-ID-
Journal

Data Mining
Kampagnen
Lifetime Value
Share of Wallet

Sammeln Das Kunden-Puzzle fügt sich zusammen



Daten-Vorwäsche

erfolgt unter Nutzung von SAS/Base-Funktionen



Waschgang	SAS/Base-Funktionen
standardisieren	left, compl
Groß-/Kleinschreibung	substr, upcase, lowcase
Umlaute/Sonderzeichen	tranwrd (Ersetzen+Korrektur)
Textfelder splitten	scan, index, substr
Vollständigkeit von Feldern, Anzahl Vokale/Konsonanten	substr
Feldabhängigkeiten, z.B. Geschlecht-Vorname, PLZ-Ort (Länderformate)	spedis
Rechtschreibkorrektur, Fantasienamen, Text Mining	spedis



Namen	Straße-Hausnummer	PLZ	Ort	Geburtsdatum
Schäfer, Manfred	Blumenstr. 7	50678	Köln	.
Schäfer, Manfred	Blumenweg 7	50678	Köln	05.01.19 67
Schäfer, M.	Blumenstr. 7a	50678	Köln	05.01.19 76

Deduplikation

Spedis-Algorithmus klassifiziert nach Ähnlichkeit

Ähnlichkeits-Klasse	Spedis-Intervall ($0 < S1 < S2 < 200$)	Aktion	Vorteil
Match-Group gleich oder sehr ähnlich	$< S1$	“Dublette!“	Zeitgewinn, Entlastung
Match-Control-Group Ähnlichkeit unsicher	$\geq S1, \leq S2$	Verifizieren	Risikominderung von Fehldeduplikation
Non-Match-Group ungleich oder sehr unähnlich	$> S2$	„Neukunde!“	Zeitgewinn, Entlastung

Die Spedis-Funktion von SAS bringt schnellen Nutzen statt langfristige Kosten!

Check-Point	Query	Keyword	SPEDIS-Wert
match	fuzzy	fuzzy	0
singlet	fuzy	fuzzy	6
doublet	fuuzzy	fuzzy	8
swap	fzuzy	fuzzy	10
truncate	fuzz	fuzzy	12
append	fuzzys	fuzzy	5
delete	fzzy	fuzzy	12
insert	fluzzy	fuzzy	16
replace	fizzy	fuzzy	20
firstdel	uzzy	fuzzy	25
firstins	pfuzzy	fuzzy	33
firstrep	wuzzy	fuzzy	40
several	floozy	fuzzy	50

- Wertebereich 0-200 (0=match)
- asymmetrisch->Mittelwert bilden
- Datumsformate in Datumstext, sonstige numerische Formate in Text konvertieren
- Mehrfeldervergleich über vorheriges Konkattenieren

Fehl-Mapping	Lösung mit Eintrag in Id-Journal-Tabelle
n:1- Overmapping	n:m mit $n \geq m > 1$
n:m- Undermapping mit $n \geq m > 1$	n:m ₀ mit $m_0 < m$

Thomas Rüdiger
thomas.ruediger@gmx.at

**Deduplication
made in Germany**

02/2002

Einführung von SAS-Deduplikation im
CRM-Mart von AXA Deutschland