

# Hochschule für Technik, Wirtschaft und Kultur

Fakultät Informatik, Mathematik und Naturwissenschaften

## **Projekt zur Lehrveranstaltung Data Warehousing**

Betriebswirtschaftliche Analyse und Entwurf des Data Warehouse

**Datum:** 29. Juni 2015

**Verfasser:** Paul Jaehne und Henry Glück

## Inhaltsverzeichnis

<b>1</b>	<b>Betriebswirtschaftliche Analyse</b>	<b>3</b>
1.1	Beschreibung möglicher Anwendungen aus Business-Sicht . . . . .	3
1.2	Konzeptuelle Modellierung . . . . .	4
1.3	Datenverarbeitungsanforderungen . . . . .	7
<b>2</b>	<b>Entwurf des Data Warehouse</b>	<b>8</b>
2.1	Relationale Umsetzung eines MDM-Schemas . . . . .	8
2.2	Optimierung der Data Cubes . . . . .	9

# 1 Betriebswirtschaftliche Analyse

## 1.1 Beschreibung möglicher Anwendungen aus Business-Sicht

Um Faktoren und Kennzahlen für den langfristigen Erfolg des Unternehmens zu erstellen, können die bereitgestellten Daten des Online-Shops aus verschiedenen Perspektiven betrachtet werden. Ein mögliches Vorgehensmodell bildet hierfür die Balanced-Scorecard.

Die bedeutendste Perspektive bei Analyse der bereitgestellten Daten ist die finanzielle Sicht. Zu dieser gehören typische betriebswirtschaftliche Kennzahlen wie der Gewinn, der Umsatz und die Rentabilität. Als Ziel für diese Perspektive kann die Steigerung der genannten Kennzahlen angesehen werden. Auch die Identifikation und Abstoßung unrentabler Produkte kann als Ziel für die finanzielle Perspektive betrachtet werden. Da jedoch nur die Verkaufspreise zur Verfügung stehen, beschränkt sich die Analyse auf die Umsatzentwicklung und -verteilung. Der Umsatz soll dabei in Abhängigkeit der Dimensionen Zeit, Produkt, Alter bzw. Geschlecht der Kunden betrachtet werden. Auf Basis der Ergebnisse dieser Analyse lassen sich beispielsweise die Top-Seller oder auch die kaufkräftigsten Kundengruppen bestimmen.

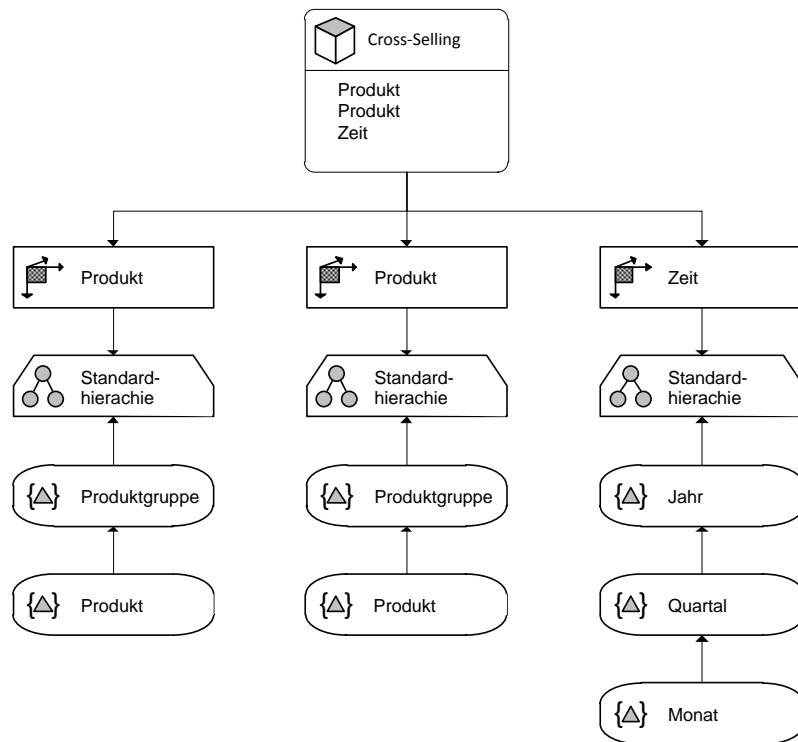
Eine weitere Betrachtungsweise ist die Kundenperspektive. Mögliche Ziele sind hierbei eine hohe Kundenzufriedenheit sowie ein hoher Neukundenanteil. Eine denkbare Fragestellung wäre auch, welche Produkte werden von Kunden häufig zusammen gekauft, was auch als Cross-Selling bezeichnet wird. Mithilfe dieser Analyse können für potentielle Kunden spezielle Angebote erstellt werden, um so den Umsatz pro Bestellung zu erhöhen. Daraus abgeleitete Kennzahlen sind die Bestellungen pro Kunde und Artikel pro Bestellung.

Zur Analyse und Bewertung der internen Geschäftsprozesse können die Bearbeitungszeiten der Bestellungen oder die Retouren betrachtet werden. Diese gilt es zu verringern. So könnte bei der Anzahl der Retouren analysiert werden, ob es bei diesen einen besonders häufigen Grund gibt. Sollten bestimmte Artikel beispielsweise besonders häufig wegen mangelnder Qualität reklamiert werden, so liegt die Empfehlung nahe, diese aus dem Warenangebot zu entfernen.

Der vierte Punkt der Balanced-Scorecard wird als Lernen und Wachstum bezeichnet. Dabei geht es um die Zufriedenheit der Mitarbeiter und deren Qualifikationen. Dazu sind jedoch keine Daten vorhanden und eine Analyse daher nicht möglich.

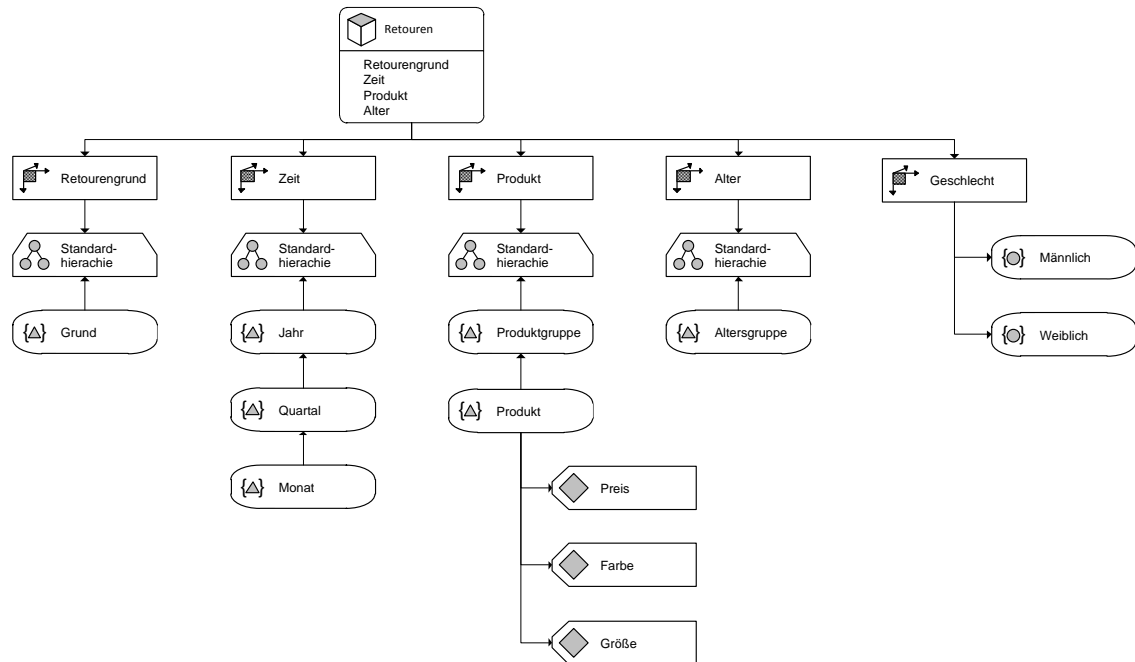
## 1.2 Konzeptuelle Modellierung

### Cross-Selling



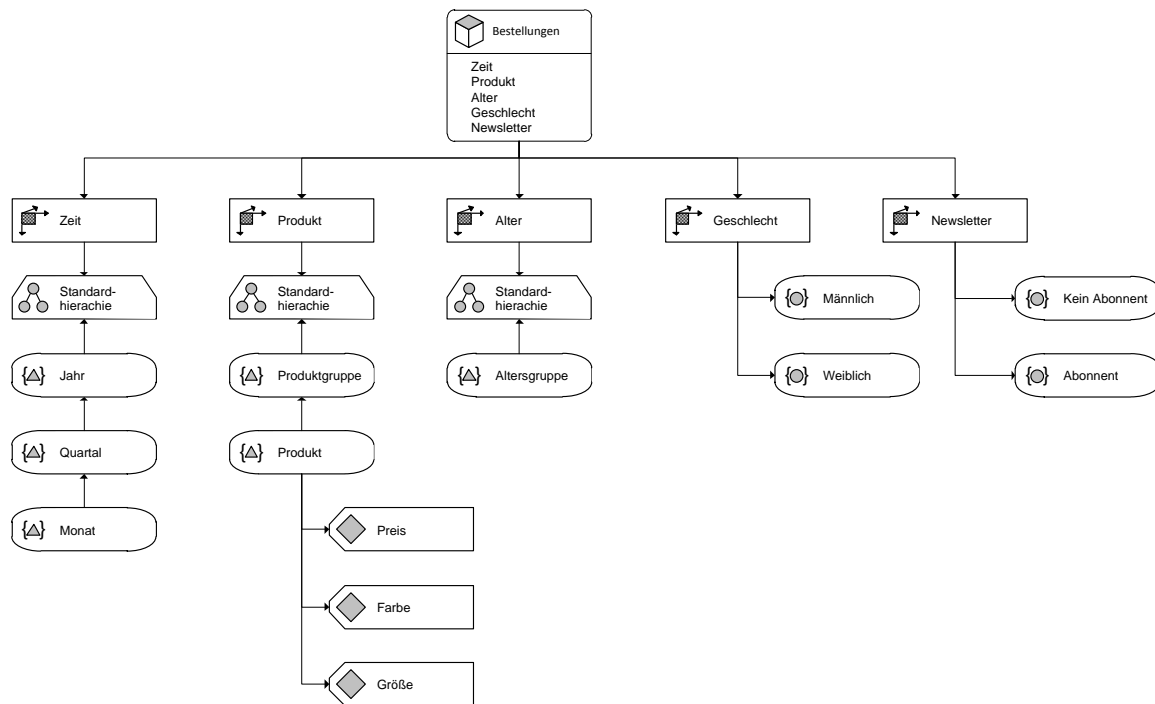
Mithilfe dieses Cubes soll ermittelt werden, welche Produkte bzw. Produktgruppen besonders häufig zusammen gekauft werden. Auf Basis der Ergebnisse dieser Untersuchung können beispielsweise besondere Angebote oder Produktkombinationen erstellt werden, um so den Umsatz pro Bestellung zu erhöhen.

## Retouren



Der Cube *Retouren* beinhaltet die Anzahl der Retouren verteilt auf die Dimensionen *Rückgabegrund*, *Zeit*, *Produkt*, *Alter* bzw. *Geschlecht* der Kunden. Darüber soll untersucht werden, welche Gründe häufig zur Retoure führen, wann gehäuft Retouren auftreten, welche Produkte zurückgegeben werden und wie sich verschiedene Kundengruppen verhalten. So können unter Umständen häufig reklamierte Produkte erkannt und aus dem Angebot entfernt werden.

## Bestellungen



Mit dem Cube *Bestellungen* soll die Anzahl der Bestellungen in Abhängigkeit der *Zeit*, des *Produktes* sowie der Kundenattribute *Alter*, *Geschlecht* und *Newsletter-Abonement* betrachtet werden. Mithilfe dieser Analyse lassen sich sowohl die „Top-Seller“ als auch die „Ladenhütter“ des Sortiments auf einfache Weise erkennen. Zusammen mit dem Verkaufspreis und der Anzahl der Retouren aus der vorherigen Untersuchung lässt sich so auch die Umsatzverteilung bezogen auf die einzelnen Produkte und Kundengruppen berechnen. Durch Dimension *Newsletter* soll ermittelt werden, ob dieser einen Einfluss auf Bestellverhalten der einzelnen Kundengruppen hat.

### **1.3 Datenverarbeitungsanforderungen**

Da als Datenquelle die operative Datenbank des Online-Shop dient, stellt diese bezogen auf die Granularität, die Aktualität und die Glaubwürdigkeit eine ausreichend hohe Datenqualität bereit. Für weiterführende Analysen bezüglich des Gewinns oder der Rentabilität sind die Daten der Datenbank jedoch nicht ausreichend, da Informationen über Einkaufspreise und Ähnliches nicht zur Verfügung stehen. Hierfür müssten externe Datenquellen, wie Preislisten von Großhändlern, in den Analyseprozess integriert werden.

Für das Auslesen der Basisdaten bietet sich eine periodische Extraktion an, da der Fokus bei den Analysen auf der zeitlichen Entwicklung bzw. Verteilung der festgelegten Kennzahlen liegt. Da in der Zeit-Dimension der Monat die kleinste Konsolidierungsstufe bildet, genügt es die Datenextraktion und Auswertung ein- bis zweimal im Monat durchzuführen.

## 2 Entwurf des Data Warehouse

### 2.1 Relationale Umsetzung eines MDM-Schemas

Zur Realisierung, der in Phase 1 entworfenen Data Cubes, sollen die nachfolgenden relationalen Modelle verwendet werden. Es wurde sich hierbei für das Star-Schema entschieden. Dieses bietet gegenüber dem Snowflake-Schema eine einfachere Struktur der Cubes und eine effizientere Anfrageverarbeitung durch die Vermeidung von JOIN-Operationen. Die notwendigen Befehle zum Anlegen der Dimensions- und Faktentabelle sind in der beigefügten SQL-Datei zu finden.

#### 2.1.1 Data Cube für die Anzahl der Bestellungen

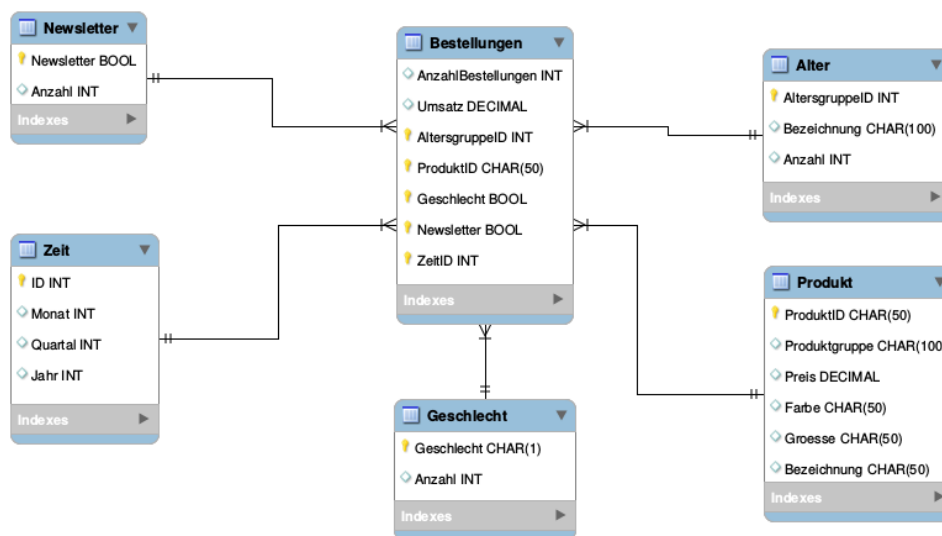


Abbildung 1: Relationales Umsetzung des Cubes *Bestellungen*

Der Data Cube *Bestellungen* setzt sich aus der Faktentabelle *Bestellungen* und den Dimensionstabellen *Alter*, *Produkt*, *Geschlecht* und *Zeit* zusammen.



### 2.1.2 Data Cube für die Anzahl der Retouren

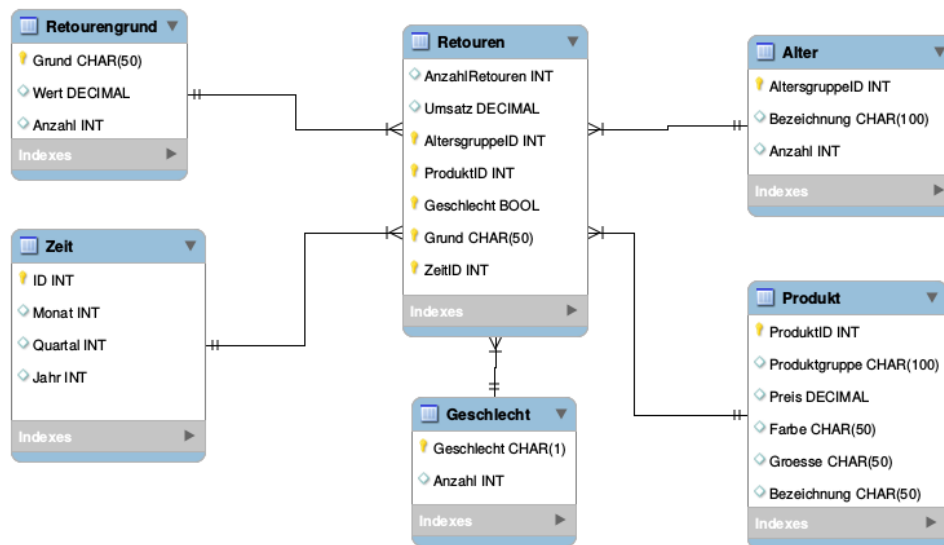


Abbildung 2: Relationales Umsetzung des Cubes *Retouren*

Analog zum Cube *Bestellungen* wurde auch der Data Cube *Retouren* zur Analyse der Reklamationen in eine Faktentabelle und mehrere Dimensionstabellen untergliedert.

### 2.1.3 Data Cube für Cross-Sells

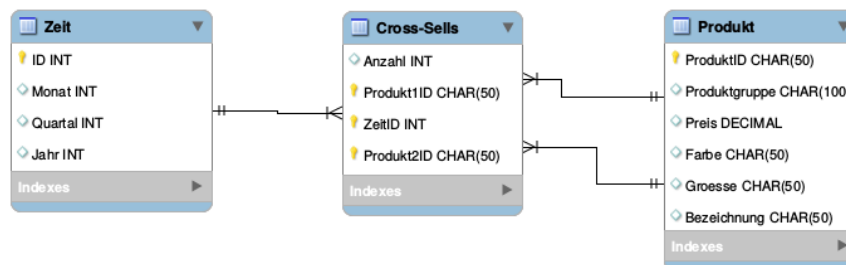


Abbildung 3: Relationales Umsetzung des Cubes *Cross-Sells*

Mithilfe des Data Cubes *Cross-Sells* soll, wie bereits erwähnt, untersucht werden, welche Produkte besonders häufig zusammen gekauft werden. Da zwischen den Produkten eine Art m-zu-n-Beziehung besteht, besitzt die Faktentabelle *Cross-Sells* zwei Fremdschlüssel auf die Dimensionstabelle *Produkt*.

## 2.2 Optimierung der Data Cubes

Zur Optimierung der Data Cubes stehen verschiedene Möglichkeiten zur Verfügung. So könnten durch materialisierte Sichten sowohl für die Cubes also auch für die Basis-Datenbank das Laden der Daten bzw. der Analyseprozess beschleunigt werden. Dabei

wäre es ausreichend, wenn diese wöchentlich aktualisiert werden. Auch die Verwendung von Indexen wäre denkbar, um so die Abarbeitung bestimmter Analyseabfragen zu beschleunigen.