

Inhaltsverzeichnis

Abbildungsverzeichnis	ix
Tabellenverzeichnis	xiii
Symbolverzeichnis	xv
Abkürzungsverzeichnis	xxi
1 Einleitung	1
1.1 Motivation	1
1.2 Aufbau und Inhalt der Arbeit	2
I Integration verteilter Datenbestände zum Aufbau entscheidungsunterstützender Datenbanksysteme	5
2 Datenbanksysteme zur Entscheidungsunterstützung	7
2.1 Einführung in das Data Warehouse-Konzept	7
2.1.1 Grundlagen	7
2.1.2 Modellierung	12

2.1.3	Architektur	17
2.2	Anwendungen entscheidungsunterstützender Informationssysteme	19
2.2.1	Ad-hoc-Anfragen	19
2.2.2	Wiederkehrende Analysen	20
2.2.3	Data Mining	21
2.2.4	Weitere Anforderungen	28
2.3	Virtuelle Data Warehouses	29
2.3.1	Definition	29
2.3.2	Verteilte Datenbanksysteme	29
2.3.3	Einsatzszenarien	30
2.3.4	Replikation	31
3	Anfrageverarbeitung in klassischen Datenbanksystemen	35
3.1	Grundlagen und Begriffsbestimmungen	35
3.1.1	Phasen der Generierung von Ausführungsplänen	35
3.1.2	Anfrageformalismen am Beispiel der relationalen Algebra	37
3.1.3	Zugriffsstrukturen	41
3.1.4	Operatoren und Ausführungspläne	49
3.2	Kostenbasierte Optimierung	56
3.2.1	Grundlagen	56
3.2.2	Kostenschätzung	59
3.2.3	Optimierung von Anfrageplänen	67
3.3	Parallelverarbeitung	80
3.3.1	Parallele Architekturen	80
3.3.2	Optimierung in parallelen Systemen	80
4	Anfrageverarbeitung in föderierten Datenbanksystemen	83
4.1	Grundlagen und Begriffsbestimmungen	84

4.2	Eigenschaften	86
4.2.1	Verteilung	86
4.2.2	Heterogenität	87
4.2.3	Autonomie	88
4.3	Kostenbasierte Optimierung	89
4.3.1	Grundlagen	89
4.3.2	Generierung alternativer Pläne am Beispiel von Aufzählungsregeln	91
4.3.3	Kostenbasierte Auswahl des auszuführenden Plans	100
4.3.4	Verwendung mehrerer Threads zur Ausführung	104
II	Anfrageverarbeitung in virtuellen Data Warehouses	107
5	Anforderungen an die Anfrageoptimierung in virtuellen Data Warehouses	109
5.1	Anforderungen an ein virtuelles Data Warehouse	109
5.2	Anfrageverarbeitung	113
6	Multidimensionale Erweiterung der relationalen Algebra	117
6.1	Grundlagen	118
6.1.1	Anforderungen an eine multidimensionale Algebra	118
6.1.2	Beispiel eines Data Warehouses	120
6.2	Konzepte der multidimensionalen Algebra	121
6.2.1	Elemente von Datenwürfeln	121
6.2.2	Datenwürfel-Schema	133
6.2.3	Datenwürfel-Instanz	135
6.3	Operationen der multidimensionalen Algebra	136
6.3.1	Übertragung von relationalen Operationen	136

6.3.2 OLAP-Operationen und ihre Umsetzung in der multidimensionalen Algebra	157
7 Globale Anfrageoptimierung	161
7.1 Kostenschätzung für multidimensionale Operationen	161
7.1.1 Abgrenzung	161
7.1.2 Einstellige Operationen	163
7.1.3 Zweistellige Operationen	166
7.2 Erzeugung und Bewertung von Ausführungsplänen	168
7.2.1 Erzeugung alternativer Pläne	168
7.2.2 Zuordnung von Teilanfragen bei replizierten Daten	176
7.3 Aufzählung, Auswahl und Anpassung der auszuführenden Pläne	182
7.3.1 Bestimmung eines kostenoptimalen Plans	182
7.3.2 Materialisierte Sichten und Caching	188
7.3.3 Adaptive Anfrageverarbeitung	192
7.4 Optimierung von relationalen OLAP-Anfragen	199
7.4.1 Grundlagen	199
7.4.2 Simultane Optimierung bei Einschränkung mehrerer Dimensionen	200
7.4.3 Verwendung transienter Sichten	203
III Umsetzung einer Testumgebung und Bewertung der erzielten Ergebnisse	211
8 Implementierungsalternativen	213
8.1 Anforderungen an die Föderierungsebene eines virtuellen Data Warehouses	213

8.2 Alternative Entwicklungsansätze	214
8.2.1 Grundsätzliche Architekturen	214
8.2.2 Vollständige Neuentwicklung	218
8.2.3 Einsatz von Standardsoftware	219
8.3 Umsetzung des Prototyps	220
8.3.1 Ziele der prototypischen Implementierung	220
8.3.2 Verwendete Komponenten	221
8.3.3 Verknüpfung der Komponenten	223
9 Evaluierung des Laufzeitverhaltens multidimensionaler Operationen	227
9.1 Ausgestaltung der Testumgebung zur Evaluierung	228
9.1.1 Das Data Mart ‘Kosten und Provisionen’	228
9.1.2 Das Data Mart ‘Aktiv und Passiv’	229
9.2 Auswertung der Testergebnisse	230
9.2.1 Grundlagen	230
9.2.2 Push-Down-Fact-Scan	233
9.2.3 Selektion	242
9.2.4 Vereinigung	246
9.2.5 Verbund	250
9.3 Bewertung der Resultate	252
10 Schlußbetrachtung	255
10.1 Zusammenfassung und Ergebnisse	255
10.2 Weitere Forschungsgebiete im Kontext dieser Arbeit	258
10.2.1 Alternative Ansätze zum Design von Informationssystemen	258
10.2.2 Neue Herausforderungen	265
10.2.3 Offene Problemfelder	266
10.3 Ausblick	268

Anhang	269
A Beispieldaten	271
A.1 Kennzahlendimension	271
A.2 Dimensionen	272
A.3 Kennzahlen	272
A.4 Hierarchien	274
A.5 Dimensionsattribute	277
Literaturverzeichnis	281
Thesen	317