Informatik

Arkadiy Omelchenko

Hierarchische physische Data-Cube-Strukturen in einem mobilen Data-Warehouse

Diplomarbeit

Diplom.de

Bibliografische Information der Deutschen Nationalbibliothek:

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de/ abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlages. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2007 Diplom.de
ISBN: 9783836605427

https://www.diplom.de/document/229009

Arkadiy Omelchenko

Hierarchische physische Data-Cube-Strukturen in einem mobilen Data-Warehouse

Diplom.de

Diplomarbeit

ISBN: 978-3-8366-0542-7

Arkadiy Omelchenko

Hierarchische physische Data-Cube-Strukturen in einem mobilen Data-Warehouse

Diplom.de

Arkadiy Omelchenko
Hierarchische physische Data-Cube-Strukturen in einem mobilen Data-Warehouse

ISBN: 978-3-8366-0542-7
Druck Diplomica® Verlag GmbH, Hamburg, 2007
Zugl. Freie Universität Berlin, Berlin, Deutschland, Diplomarbeit, 2007

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtes.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Die Informationen in diesem Werk wurden mit Sorgfalt erarbeitet. Dennoch können Fehler nicht vollständig ausgeschlossen werden, und die Diplomarbeiten Agentur, die Autoren oder Übersetzer übernehmen keine juristische Verantwortung oder irgendeine Haftung für evtl. verbliebene fehlerhafte Angaben und deren Folgen.

© Diplomica Verlag GmbH
http://www.diplom.de, Hamburg 2007
Printed in Germany

Kurzfassung

Die Entscheidungsunterstützung in den modernen Unternehmen mit ihren immensen Informationsmengen ist ohne das Konzept eines *Data-Warehouse* und *OLAP* nicht mehr vorstellbar. Außerdem werden die mobilen Geräte wie PDAs oder Notebooks mit ihren Möglichkeiten der drahtlosen Kommunikation immer beliebter in dem unternehmerischen Alltag. Daher befassen wir uns in unserer Arbeit ausführlich mit dem *mobile O-LAP* (kurz *mOLAP*), wo die mobilen Klienten die OLAP-Dienste durch eine drahtlose Verbindung in Anspruch nehmen, sowie mit den für die Übertragung eingesetzten physischen Data-Cube-Datenstrukturen. Wir entwickelten dabei eine neue Datenstruktur *m-Dwarf,* die speziell für den mOLAP konzipiert ist. Sie speichert keine Aggregatinformationen und ist somit sehr kompakt. Sie ist um 30-35 % kleiner als fcg-Dwarfs und um 40% kleiner als Summary Tables im Durchschnitt über ein hDCL. Mit dem Einsatz der m-Dwarfs haben wir den besten existierenden Scheduling-Algorithmus h-FCLOS wesentlich verbessert. Wir haben diesem neuen Algorithmus den Namen $FCLOS_{mD}$ (*FCLOS with m-Dwarfs*) gegeben. Mit $FCLOS_{mD}$ wird im mOLAP die durchschnittliche Wartezeit um 40%, der durchschnittliche Energieverbrauch um 40% und das generierte Datenvolumen um 30% verringert.

Schlagwörter: OLAP, mobile OLAP, mOLAP, dwarf, mobile dwarf, m-dwarf, FCLOS, h-FCLOS, $FCLOS_{mD}$

Abstract

The decision support in the modern businesses with their immense information amount is not more imaginable without the concept of the *data warehouse* and *OLAP*. The mobile devices like PDAs or notebooks are becoming more and more popular in the business workaday usage. That why we concern ourselves in our work in detail with the *mobile OLAP (mOLAP* for short*),* where the mobile clients take the OLAP services over the wireless connection, and with the transmitting of the used physical data cube structures. Thereby we developed one new data structure *m-Dwarf* that is especially designed for mOLAP. It saves no aggregate information and thus is very compact. It is 30-35% smaller than the fcg-Dwarfs and 40% smaller than the summary tables in average over an hDCL. By the m-Dwarfs usage we substantially improved the best scheduling algorithm h-FCLOS. We gave the name $FCLOS_{mD}$ to this new algorithm (*FCLOS with m-Dwarfs*). With $FCLOS_{mD}$ the average access time is decreased at 40%, the average energy consumption at 40% and the generated traffic at 30% in the mOLAP.

Keywords: OLAP, mobile OLAP, mOLAP, dwarf, mobile dwarf, m-dwarf, FCLOS, h-FCLOS, $FCLOS_{mD}$

Inhaltsverzeichnis

Kurzfassung **1**

Abstract **1**

Inhaltsverzeichnis **3**

Abbildungsverzeichnis **6**

Tabellenverzeichnis **7**

Vorwort **8**

1 Einführung **9**

1.1 Grundlagen 9

1.2 Problembeschreibung 9

1.3 Unsere Ziele und unser Beitrag 10

1.4 Überblick 11

2 Grundlagen von Data-Warehousing, OLAP und mOLAP **13**

2.1 Data-Warehousing 13

2.1.1 Einführung 13

2.1.2 Terminologie und Definitionen 14

2.1.3 Architektur eines Data-Warehouse 14

2.1.4 Datenmodellierung für Data-Warehouses 16

2.1.5 ROLAP and MOLAP 18

2.2 OLAP 18

2.2.1 Data-Cube und der cube-Operator 19

2.2.2 DCL und hDCL 23

2.3 mOLAP 27

2.3.1 Architektur 27

2.3.2 Scheduling-Algorithmen 29

2.3.3 FCLOS und h-FCLOS 31

3 Physische Data-Cube-Datenstrukturen **34**

3.1 Stand der Dinge und Related Work 34

3.1.1 ROLAP 34

3.1.2 MOLAP 35

3.1.3 Index-Strukturen 35

3.1.4 View Materialisierung 38

3.1.5 Data-Cube-Datenstrukturen 39

3.2 Dwarf-Technologie ... 44
3.2.1 Dwarf ... 44
3.2.2 Hierarchical Dwarf (h-Dwarf) ... 51
3.2.3 Coarse-grained Dwarf (cg-Dwarf) und Fully coarse-grained Dwarf (fcg-Dwarf) ... 52
3.2.4 Persistenz von Dwarfs, h-Dwarfs und cg-Dwarfs ... 54

4 Mobile Dwarf (m-Dwarf) ... 56

4.1 Motivation ... 56
4.2 Kompromiss in mOLAP ... 56
4.3 Struktur des m-Dwarf ... 58
4.4 Erzeugen des m-Dwarf ... 61
4.4.1 *CreateRMDwarf* Algorithmus und rm-Dwarf ... 61
4.4.2 *CreateMDwarf* Algorithmus ... 62
4.5 Persistenz von m-Dwarfs ... 63

5 Implementierung ... 64

5.1 Einführung und Programmdesign ... 64
5.2 Test-Datenbasis ... 68
5.3 Implementierung von Dwarfs ... 69
5.3.1 MemoryDwarf ... 69
5.3.2 PersistentDwarf ... 71
5.4 Implementierung von h-Dwarfs ... 73
5.4.1 HierarchicalMemoryDwarf ... 73
5.4.2 HierarchicalPersistentDwarf ... 74
5.5 Implementierung von cg-Dwarfs ... 75
5.6 Implementierung von fcg-Dwarfs ... 75
5.7 Implementierung von m-Dwarfs and rm-Dwarfs ... 76

6 Test- und Simulationsergebnisse der Dwarf-Technologie und m-Dwarfs in mOLAP ... 78

6.1 Größenverhältnisse der verschiedenen Dwarfs ... 78
6.2 Simulationsergebnisse in mOLAP ... 89

7 Zusammenfassung und Ausblick ... 93

7.1 Unser Beitrag ... 93
7.2 Schlussfolgerungen ... 94
7.3 Zukünftige Aktivitäten ... 94

Literaturverzeichnis ... 95

Literatur ... 95

Patente 99

Abbildungsverzeichnis

Abbildung 2-1: Typische Architektur eines Data-Warehouse 15
Abbildung 2-2: Relationen des Sternschemas 17
Abbildung 2-3: Die Faktentabelle *Facts* 19
Abbildung 2-4: DCL für den Beispiel-Data-Cube 25
Abbildung 2-5: Zusammenhang zwischen DCL und Summary Tables 26
Abbildung 2-6: Zusammenhang zwischen hDCL und Summary Tables 27
Abbildung 2-7: mOLAP-Architektur 28
Abbildung 2-8: (h-)FCLOS Algorithmus in Aktion (für α=1) 32
Abbildung 3-1: Ein Bitmap-Index für die Spalte *Gender* in einer Datenbanktabelle 36
Abbildung 3-2: Der schematische Aufbau eines B-Baums 37
Abbildung 3-3: Ein R-Baum 37
Abbildung 3-4: Ein *Cube Tree* (in der Mitte) 39
Abbildung 3-5: BUC Prozessbaum 40
Abbildung 3-6: Ein *H-Tree* 41
Abbildung 3-7: Ein *Star-Tree* 42
Abbildung 3-8: Ein *statistic tree* 42
Abbildung 3-9: Ein *Condensed Cube* 43
Abbildung 3-10: Dwarf für die Faktentabelle *Facts* 45
Abbildung 3-11: *CreateDwarfCube* Algorithmus 48
Abbildung 3-12: *SuffixCoalesce* Algorithmus 49
Abbildung 3-13: Hierarchical Dwarf (h-Dwarf) 51
Abbildung 3-14: *Coarse-grained Dwarf* mit $G_{min} = 2$ 53
Abbildung 3-15: *Fully coarse-grained Dwarf* (mit $G_{min} = \infty$) 53
Abbildung 3-16: Konzeptionelle Persistenz eines Dwarf 54
Abbildung 3-17: Die Dwarf Datei schematisch 55
Abbildung 3-18: Die Dwarf-Datei für die Faktentabelle *Facts* 55
Abbildung 4-1: m-Dwarf ohne Optimierung von Knotenseparatoren 59
Abbildung 4-2: m-Dwarf 60
Abbildung 4-3: *CreateRMDwarf* Algorithmus 61
Abbildung 4-4: rm-Dwarf 62
Abbildung 4-5: *CreateMDwarf* Algorithmus 63
Abbildung 4-6: m-Dwarf File 63
Abbildung 5-1: Allgemeines Klassendiagramm 67
Abbildung 5-2: Verschiedene Datenbankschemas mit Beispiel-Daten 68
Abbildung 5-3: Klassendiagram für *MemoryDwarf* 70
Abbildung 5-4: Klassendiagramm für *PersistentDwarf* 72
Abbildung 5-5: Klassendiagramm für *HierarchicalMemoryDwarf* 73
Abbildung 5-6: Klassendiagramm für *HierarchicalPersistentDwarf* 74