

Informatik-Fachberichte 186

Herausgegeben von W. Brauer
im Auftrag der Gesellschaft für Informatik (GI)

Erhard Rahm

Synchronisation in Mehrrechner- Datenbanksystemen

Konzepte, Realisierungsformen und
quantitative Bewertung



Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo

Autor

Erhard Rahm
Fachbereich Informatik der Universität Kaiserslautern
Erwin-Schrödinger-Straße, 6750 Kaiserslautern
derzeitige Anschrift:
IBM T. J. Watson Research Center
P. O. Box 704, Yorktown Heights, 10598 NY, USA

CR Subject Classifications (1987): H.2.4, C.2.4, C.4, D.4

ISBN-13: 978-3-540-50348-4 e-ISBN-13: 978-3-642-74123-4
DOI: 10.1007/978-3-642-74123-4

CIP-Titelaufnahme der Deutschen Bibliothek.

Rahm, Erhard:

Synchronisation in Mehrrechner-Datenbank-Systemen: Konzepte, Realisierungsformen u. quantitative Bewertung / Erhard Rahm. – Berlin; Heidelberg; New York; London; Paris; Tokyo: Springer, 1988

(Informatik-Fachberichte; 186)

ISBN-13: 978-3-540-50348-4

NE: GT

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 vom 9. September 1965 in der Fassung vom 24. Juni 1985 zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

© by Springer-Verlag Berlin Heidelberg 1988

Vorwort

Die vorliegende Untersuchung entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Fachbereich Informatik der Universität Kaiserslautern. Dabei beschäftigte ich mich im Rahmen des von der Siemens AG finanzierten Projektes 'Mehrrechner-Datenbanksysteme' vor allem mit dem Entwurf und der quantitativen Bewertung von Synchronisationsverfahren, die zur Realisierung von Hochleistungs-Datenbanksystemen geeignet sind.

Mein besonderer Dank gilt dem Leiter des Projekts, Herrn Prof. Dr. Theo Härder, für die Anregung, mich mit dem Thema 'Synchronisation in Mehrrechner-Datenbanksystemen' zu befassen, für viele wichtige Hinweise und Anmerkungen sowie für seine ständige Diskussionsbereitschaft. Herrn Prof. Dr. Jürgen Nehmer danke ich für die Durchsicht der Arbeit sowie für hilfreiche Verbesserungsvorschläge. Wesentliche Aspekte und Anregungen ergaben sich auch in Diskussionen mit meinen Kollegen an der Universität Kaiserslautern sowie mit Herrn Dr. Horst Biller und Herrn Dr. Meinhard Köstler von der Siemens AG. Die quantitativen Bewertungen basieren weitgehend auf den Diplomarbeiten von Frau Marianna Luczak, Herrn Gerald Petry sowie Herrn Peter Scheug, in denen weit mehr als sonst üblich geleistet wurde.

Für die mühevollen Arbeit des Korrekturlesens danke ich vor allem Frau Priska Schoch sowie meinen Kollegen Herrn Volker Bohn und Herrn Thomas Wagner. Darüber hinaus gilt mein Dank Frau Andrea Krahl, die viele der Abbildungen angefertigt hat, sowie dem Regionalen Hochschulrechenzentrum Kaiserslautern, auf dessen Ressourcen ich zur Durchführung der Simulationsläufe angewiesen war.

Kaiserslautern, im Juli 1988

Erhard Rahm

Zusammenfassung

Diese Arbeit beschäftigt sich mit dem Entwurf sowie der quantitativen Analyse von Synchronisationstechniken für Mehrrechner-Datenbanksysteme mit Hinblick auf die Realisierung von Transaktionssystemen hoher Leistungsfähigkeit. Dazu werden zunächst die Anforderungen an solche Systeme näher diskutiert sowie eine Klassifikation von Mehrrechner-Datenbanksystemen vorgestellt. Zur Realisierung von Hochleistungs-Datenbanksystemen kommen dabei primär zwei allgemeine Mehrrechner-Architekturen, DB-Sharing und DB-Distribution genannt, in Betracht, wobei viele Vorteile für den DB-Sharing-Ansatz sprechen. Daher konzentrieren wir uns auch auf die Synchronisation in DB-Sharing-Systemen, welche dort eine leistungsbestimmende Funktion ausübt.

Dennoch wird in einem eigenen Teil relativ ausführlich auf das umfangreiche Feld der Synchronisation in zentralisierten und verteilten Datenbanksystemen eingegangen. Zur Synchronisation bei DB-Sharing wird dann - unter Berücksichtigung bereits vorgeschlagener oder implementierter Verfahren - ein breites Lösungsspektrum untersucht. Betrachtet werden vor allem Sperrverfahren und optimistische Protokolle, die jeweils unter zentraler oder verteilter Kontrolle realisiert werden können. Als Entwurfsalternativen bzw. Optimierungsmöglichkeiten werden u.a. Mehrversionen-Konzepte, Hardware-Lösungen, Spezialprotokolle für High-Traffic-Objekte und eine Beschränkung auf Konsistenzebene 2 in die Überlegungen einbezogen. Der Entwurf neuer Synchronisationstechniken erfolgt stets unter Beachtung der engen Verflechtungen zu den anderen Systemfunktionen (Systempufferverwaltung, Lastkontrolle, Logging und Recovery), wobei aus Leistungsgründen insbesondere für das durch die DB-Sharing-Architektur eingeführte Veralterungsproblem integrierte Lösungen entwickelt werden. Dies gilt in besonderem Maße für das vielversprechende Primary-Copy-Sperrverfahren, dessen Realisierung und Zusammenwirken mit anderen Systemfunktionen ausführlich dargestellt werden.

Bei der quantitativen Bewertung der Synchronisationsprotokolle wird zunächst ein Überblick über bisherige Leistungsanalysen in zentralisierten DBS als auch in Mehrrechner-DBS gegeben. Der Schwerpunkt liegt dann auf der Beschreibung eines eigens entwickelten, detaillierten Simulationssystems sowie den damit erzielten Resultaten. In diesem System wurden neben sechs Synchronisationsverfahren für DB-Sharing auch weitere Systemkomponenten wie die Systempufferverwaltung und die Logging-Komponente vollständig implementiert; ein hoher Realitätsbezug wird zudem durch den gewählten Ansatz der Trace-getriebenen Simulation gewährleistet. Dazu standen sechs empirische Lasten aus realen DB/DC-Anwendungen zur Verfügung, welche zur Verarbeitung unter Einsatz einer Routing-Strategie auf die Rechner des simulierten DB-Sharing-Systems aufgeteilt werden. Ausführlich analysiert werden vor allem die aufschlußreichen Simulationsergebnisse für das Primary-Copy-Sperrverfahren sowie für optimistische Protokolle, die auf einer Token-Ring-Topologie basieren.

Inhaltsverzeichnis

I Einführung	1
1. Einleitung	1
2. Mehrrechner-Datenbanksysteme	6
2.1 Klassifikation von Mehrrechner-DBS	6
2.2 DB-Sharing vs. DB-Distribution	9
2.3 Schlüsselkonzepte zur Realisierung von Hochleistungs-DBS	14
2.3.1 Techniken zur Erlangung hoher TA-Raten und kurzer Antwortzeiten.....	14
2.3.2 Fehlertoleranz-Konzepte	16
II Allgemeine Synchronisationskonzepte in zentralisierten und verteilten Datenbanksystemen	20
3. Grundlagen	20
3.1 Korrektheitskriterium der Synchronisation	21
3.2 Anforderungen an die Synchronisationskomponente	24
4. Synchronisationstechniken für zentralisierte DBS	27
4.1 Sperrverfahren	28
4.1.1 RX-, RAX-, RAC-Sperrverfahren	28
4.1.2 Deadlock-Behandlung	30
4.1.3 Synchronisation auf High-Traffic-Objekten	32
4.2 Optimistische Synchronisationsverfahren	33
4.2.1 BOCC und FOCC	34
4.2.2 Das BOCC+-Verfahren	36
4.2.3 Kombination mit Sperrverfahren	38
4.2.4 Konsistenzebene 2 mit optimistischer Synchronisation	40
4.3 Zeitmarkenverfahren	41
4.4 Mehrversionen-Konzept	42
4.5 Bewertung der vorgestellten Synchronisationskonzepte	42
5. Synchronisation in verteilten Datenbanksystemen	46
5.1 Synchronisationsverfahren bei fehlender Datenreplikation	48
5.1.1 Sperrverfahren in VDBS	48
5.1.2 Optimistische Protokolle in VDBS	51
5.1.3 Zeitmarkenverfahren in VDBS	55
5.2 Update-Problematik bei replizierten Datenbanken	56

III Synchronisationsverfahren für DB-Sharing-Systeme	60
6. Systemmodell und funktionelle Komponenten	60
6.1 Lastkontrolle bei DB-Sharing	63
6.2 Logging und Recovery	67
6.3 Das Veralterungsproblem	68
6.4 Synchronisationsproblematik bei DB-Sharing	71
6.5 Abhängigkeiten funktioneller Komponenten in DB-Sharing-Systemen	74
7. Das Primary-Copy-Sperrverfahren	77
7.1 Basisverfahren	77
7.2 Optimierte Synchronisierung von Lesezugriffen	83
7.3 Integrierte Lösung des Veralterungsproblems	89
7.4 Kooperation mit der Lastkontrolle	98
7.5 Behandlung von Rechnerausfällen	102
8. Weitere Sperrverfahren für DB-Sharing	110
8.1 Zentrale Sperrverfahren	110
8.2 Synchronisation bei den DEC VAX-Clustern und bei AIM/SRCF	116
8.3 Pass-the-Buck-Protokolle	117
8.3.1 Synchronisation bei IMS Data Sharing	117
8.3.2 Erweiterungen des Pass-the-Buck-Protokolls	119
9. Optimistische Synchronisationsverfahren für DB-Sharing	124
9.1 Optimistische Protokolle mit zentraler Validierung	124
9.2 Optimistische Protokolle mit verteilter Validierung	129
9.2.1 Token-Ring-Ansatz	129
9.2.2 Broadcast-Validierung	131
9.2.3 Primary-Copy-artige Synchronisation	134
9.2.3.1 Broadcast-Validierung mit Primary-Copy-artiger Synchronisation	135
9.2.3.2 Kombination mit dem Primary-Copy-Sperrverfahren.....	138
9.3 Abschätzung des Validierungs- und Kommunikationsaufwandes	139
10. Entwurfsalternativen und Optimierungsmöglichkeiten	143
10.1 Asynchrone Sperranforderungen	143
10.2 Mehrversionen-Konzept bei DB-Sharing	144
10.3 Beschränkung auf Konsistenzebene 2	148
10.4 Einsatz einer nahen Rechnerkopplung	149
10.5 Synchronisation auf Eintrags- und Satzebene	152
10.6 Synchronisation von High-Traffic-Objekten bei DB-Sharing	155
11. Vergleich der vorgestellten Konzepte und Verfahren	157

IV	Quantitative Leistungsuntersuchungen	166
12.	Existierende Leistungsanalysen zur Synchronisation in DBS	166
12.1	Methoden der quantitativen Leistungsbewertung von Synchronisationsverfahren	166
12.2	Existierende Leistungsanalysen für zentralisierte und verteilte DBS	168
12.3	Existierende Leistungsanalysen für DB-Sharing	171
12.3.1	Arbeiten am IBM Research Center in Yorktown Heights	171
12.3.2	Simulationsarbeiten an der Universität Stuttgart	174
12.3.3	Bisherige empirische Simulationen an der Universität Kaiserslautern	176
13.	Beschreibung des implementierten Simulationssystems sowie der verwendeten Referenz-Strings	179
13.1	Aufbau der Referenz-Strings	180
13.2	Aufbau, Parametrisierung und Realisierung des Simulationssystems	181
13.2.1	Grobstruktur und Parametrisierung des Simulationssystems	181
13.2.2	Die Simulationssteuerung	184
13.2.3	Abarbeitung der Referenzsätze	186
13.2.4	Der Referenz-Manager	188
13.2.5	Die Synchronisationskomponente	190
13.2.6	Systempufferverwaltung und Logging	192
13.2.7	Das Kommunikationssystem	194
13.3	Beschreibung der Transaktionslasten	195
13.4	Bestimmung der Routing-Tabellen und der PCA-Verteilungen	197
13.5	Ergebnisgrößen und Ausgabedaten	200
14.	Quantitative Bewertung des Primary-Copy-Sperrverfahrens	202
14.1	Einfluß der Synchronisation auf FPA/DBTT-Seiten	202
14.1.1	Durchsatzverhalten	202
14.1.2	CPU-Auslastung	207
14.1.3	Antwortzeitverhalten	212
14.1.4	E/A-Verhalten	217
14.1.5	Sperr- und Kommunikationsverhalten	219
14.2	Einfluß der Lastverteilung	226
14.3	Einfluß der Kommunikationskosten	229
14.4	Zusammenfassende Bewertung der Simulationsresultate	234
15.	Quantitative Analyse weiterer Synchronisationsverfahren für DB-Sharing	237
15.1	Simulationsergebnisse für die Token-Ring-Protokolle	237
15.1.1	Ergebnisse bei TER	237
15.1.2	Ergebnisse bei DOD	246
15.1.3	Zusammenfassende Bewertung	250
15.2	Beurteilung der Synchronisationsverfahren unter zentraler Kontrolle	253
V	Resümee und Ausblick	256
VI	Literatur	262