

# Die fünf großen W

## (Warum – Wann – Was – Wo - Wie)

### der Datenbankreorganisation bei relationalen Datenbank-Management-Systemen

*Stefan Dorendorf*

Friedrich-Schiller-Universität Jena

Institut für Informatik

Lehrstuhl für Datenbanken und Informationssysteme

Ernst-Abbe-Platz 1-4

07743 Jena

[stefan@ba-glauchau.de](mailto:stefan@ba-glauchau.de)

#### 1 WARUM - Gründe für eine Datenbankreorganisation

Die Gründe, die zur Durchführung einer Datenbankreorganisation führen, sind verschieden. So kann es im Laufe der Nutzung einer Datenbank, abhängig vom Datenbank-Management-System (DBMS) und der Nutzungsart, zu einer schleichenden Leistungsver schlechterung kommen oder das System stößt an Grenzen, die bei der Implementierung festgelegt wurden. Es wird nach einer gewissen Nutzungszeit notwendig, diese Probleme zu beseitigen. In dem Zusammenhang spricht man häufig von einer Datenbankreorganisation.

Ein Grund für die Leistungsver schlechterung können Degenerierungen der zur Datenspeicherung und zum Wiederauffinden der Daten verwendeten Strukturen sein, wenn diese aus Performancegründen bei Änderungsoperationen nicht vollständig oder nicht sofort gepflegt werden. So werden bei vielen Systemen bei Einfüge- und Änderungsoperationen Sortierreihenfolgen innerhalb der Datenbestände nicht beachtet oder es kommt zur Fragmentierung des Datenbestands. Bei tupelverlängernden Änderungsoperationen kann es zu Platzmangel am ursprünglichen Speicherort kommen. Betroffene Tupel werden an einen anderen Ort verschoben. Am originalen Speicherort werden Zeiger (Auslagerungszeiger) auf die jeweiligen neuen Speicherorte abgelegt.

Auch die Verteilung der Daten auf unterschiedliche Datenträger kann Einfluß auf die Performance haben. Bei einer ungünstigen Verteilung des Datenbestands werden einige Datenträger sehr stark und andere nur wenig belastet.

Die bereits erwähnten Fragmentierungen führen meist neben einer Leistungsver schlechterung oftmals noch zu einer verschlechterten Auslastung des zur Datenspeicherung belegten Sekundärspeichers.

Auch Änderungen in der Art der Nutzung des Datenbestands können eine Neuorganisation erfordern. Nach [MHR96] sind in einer operativen Datenverarbeitung mit meist kurzen Transaktionen, bei denen häufig auch Änderungen am Datenbestand vorgenommen werden, andere Organisationsformen des Datenbestands sinnvoll, als bei einem System zur Unterstützung von Entscheidungsfindungsprozessen mit meist langen und aufwendigen sequentiellen Leseprozessen.

Nach [LS93] können die Gründe, die zur Durchführung einer Datenbankreorganisation führen, in zwei Klassen eingeteilt werden: zwingende Gründe und solche, die Anpassungen empfehlenswert erscheinen lassen.

Zwingende Gründe sind z.B. das Erreichen der oben erwähnten Grenzen (Schwellwerte), welche vom Datenbank-Management-System oder vom Betriebssystem vorgegeben werden, wie z.B. maximale Dateigrößen, maximale Anzahl verwaltbarer Speichereinheiten (z.B. Extents je Relation) usw. Aber nicht nur Gründe, die in direktem Zusammenhang mit dem DBMS oder der Systemplattform stehen, sondern auch Veränderungen am betrachteten Umweltausschnitt können eine Datenbankreorganisation erzwingen. Hauptziel ist hier die Aufrechterhaltung des Betriebs des Datenbanksystems bzw. der Anwendung.

Bei einer unwirtschaftlichen Speicherplatzausnutzung oder aus Performancegründen, z.B. weil interne Sortierreihenfolgen nicht eingehalten werden und somit häufig aufwendige Sortierläufe notwendig sind, ist eine Datenbankreorganisation oft empfehlenswert. Allerdings sollte dabei der Aufwand für die Reorganisation dem zu erwartenden Nutzen gegenübergestellt werden. Abbildung 1 zeigt noch einmal eine Unterteilung der Gründe für eine Datenbankreorganisation

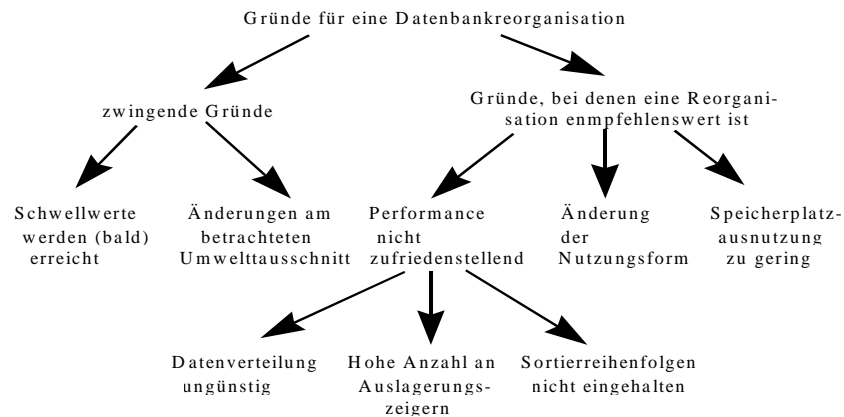


Abb. 1: Gründe für eine DB-Reorganisation

Mit einer Reorganisation können u.a. die folgenden Ziele, die teilweise einander bedingen, verfolgt werden:

- Beseitigung von Degenerierungen
- Beschleunigung der Datenzugriffe
- Anpassung der physischen Datenspeicherung an Veränderungen in der Nutzungsform des Datenbestands
- Freigabe von nicht mehr benötigtem, aber reserviertem Speicherplatz
- Erreichen eines definierten Belegungsgrads der Speicherblöcke für Daten und Indexe

## 2 WANN - sollte reorganisiert werden

Wann und wie die Reorganisation einer Datenbank durchgeführt werden kann, hängt von vielen Faktoren ab. Wichtig ist es vor allem, Aufwand und Nutzen gegeneinander abzuwägen. Einen wichtigen Faktor stellen die Reorganisationsmöglichkeiten, die das DBMS bietet, dar. Beachtung sollte dabei vor allem den möglichen Granulaten (zu reorganisierende Einheiten) für eine Datenbankreorganisation und der technischen Durchführung geschenkt werden. Eine Datenbankreorganisation, die mit hohem technischen Aufwand (Entladen der Datenbank, Zerstören und Neuanlegen der Schemaelemente, Laden der Daten) verbunden ist und darüber hinaus die Verfügbarkeit der Datenbank stark einschränkt, ist sicherlich seltener sinnvoll durchführbar als eine lokale Reorganisation einzelner Relationen.

Auch das vorhandene Schema einer Datenbank kann die Art und Häufigkeit ihrer Reorganisation beeinflussen. Vor allem die Wahl der Datentypen (z.B. Zeichenketten fester Länge oder Zeichenketten variabler Länge) hat Einfluß darauf, ob es durch Verlängerungen von Tupeln zu Auslagerungen und, damit verbunden, zu Auslagerungszeigern kommt.

Ein weiterer Faktor, der eine Datenbankreorganisation beeinflusst, ist die Nutzungsform der Datenbank. So wird eine Datenbank, die überwiegend lesend genutzt wird, wesentlich langsamer degenerieren als eine Datenbank, an der häufig Änderungen am Datenbestand vorgenommen werden.

Die Systemplattform, auf der die Datenbanklösung installiert wurde, spielt ebenfalls eine Rolle bei der Entscheidung, wann eine Reorganisation der Datenbank durchzuführen ist. Sie hat wesentlichen Einfluß auf die Ziele einer Reorganisation. Vor allem die Parallelisierung von Datenbankoperationen setzt die Unterstützung von seiten der Hardware und des Betriebssystems voraus.

Nicht zuletzt hat der Typ des zur Datenspeicherung verwendeten Speicherplatzes (Raw Device/Dateisystem/Virtual Disks) Einfluß auf Durchführung und Häufigkeit einer Datenbankreorganisation.

### 3 WAS - Granulate einer Datenbankreorganisation

Die Granulate, die für eine Datenbankreorganisation denkbar sind, lassen sich grob in zwei Gruppen einteilen. Einerseits existieren logische Granulate wie Datenbanken, Relationen oder Mengen von Relationen. Andererseits lassen sich physische Granulate bilden, die zum Teil vom verwendeten DBMS abhängig sind. Zu den physischen Granulaten zählen Partitionen, Zugriffshilfen (Indexe) und die vom Datenbank-Management-System verwendeten Speichereinheiten.

Das größte Granulat stellt **die gesamte Datenbank** dar. Damit werden alle Strukturen der Datenbank wieder bereinigt und es kann, bei ordnungsgemäßer Durchführung, mit einem unter den gegebenen Umständen „optimalen“ Ergebnis gerechnet werden. Allerdings ist eine solche Reorganisation meist auch sehr arbeits- und zeitintensiv, und ein globales Optimum im Ergebnis wird nur selten erreicht. Weiterhin sollte beachtet werden, daß hier auch Teile der Datenbank reorganisiert werden, bei denen eine Reorganisation nicht oder noch nicht erforderlich gewesen wäre.

Als weiteres Granulat für eine Datenbankreorganisation ist **die Relation** denkbar. Die Verwendung dieses Granulats erlaubt die gezielte Reorganisation degenerierter Bereiche der Datenbank. Es eignet sich besonders zum Wiederherstellen von Sortierungen innerhalb von Relationen und zur Freigabe von Speicher.

Im Falle der Nutzung von Konzepten, die eine physisch zusammenliegende Speicherung von Daten aus mehreren Relationen erlauben (Cluster-Bildung), ist als weiteres Granulat eine **Menge von Relationen** sinnvoll.

Im Zusammenhang mit Konzepten zur Partitionierung von Relationen ist, vor allem bei sehr großen Relationen, auch eine **Partition** einer Relation denkbar (vgl. auch [HR99]).

Auch **einzelne Zugriffshilfen** (Indexe) können reorganisiert werden, um zum Beispiel definierte Füllungsgrade für die Indexseiten wiederherzustellen.

Abhängig vom Speichermodell, welches das DBMS verwendet, können auch andere Granulate, die sich auf die Einheiten des Speichermodells beziehen (z.B. Table Spaces), sinnvoll sein.

### 4 WO - Reorganisationsebenen

Eine Datenbankreorganisation kann auf der Schema-, der Verteilungs- oder der internen Ebene eines Datenbanksystems durchgeführt werden.

#### ➤ Schemaebene

Als Reorganisation auf Schemaebene (Restrukturierung) soll hier eine Form der Datenbankreorganisation bezeichnet werden, bei der das Datenbankmodell der Informationsstrukturen des Umweltausschnitts verändert wird. Das kann zum Beispiel sinnvoll sein, wenn sich die Nutzungsform des Datenbestands geändert hat und das Datenbankschema an die veränderte Nutzungsform angepaßt werden soll. Da die Ursachen für eine Schemaumstellung i.d.R. außerhalb des Verantwortungsbereichs des Administrators und des verwendeten DBMS liegen, soll eine Reorganisation auf der Schemaebene zukünftig nicht weiter betrachtet werden.

#### ➤ Verteilungsebene

Bei einer Datenbankreorganisation auf der Verteilungsebene wird der Ort der physischen Abspeicherung der Daten „global“ verändert. Das heißt, es handelt sich bei einer solchen Umverteilung nicht um eine „lokale“ Umspeicherung von Daten, wie sie etwa bei Umsortierungen vorgenommen wird, sondern bspw. um eine gezielte Verteilung der Daten einer oder mehrerer Relationen auf unterschiedliche Datenträger, welche von einem Datenbank-Server verwaltet werden, um eine Lastverteilung zu erreichen. Eine solche Verteilung kann einmal dadurch

erreicht werden, daß die Relationen einer Datenbank auf unterschiedliche Datenträger verteilt werden. Auch die Trennung von Nutzerdaten, Metadaten (Datenbankkatalog) und Log-Daten erscheint oft als sinnvoll oder gar als nötig. Einige DBMS (z.B. Informix, Oracle) unterstützen eine horizontale Partitionierung, bei der die Tupel und Indexe einer Relation auf unterschiedliche Datenträger verteilt werden. Auch eine gezielte, physisch zusammenliegende Speicherung der Tupel mehrerer Relationen, die logisch in Beziehungen zueinander stehen, ist denkbar.

### ➤ Interne Ebene

Bei einer Datenbankreorganisation auf der internen Ebene steht das Beseitigen von Degenerierungen verschiedener Art im Vordergrund. Degenerierungen in den Datenbeständen und Zugriffspfaden entstehen häufig im Zusammenhang mit Änderungsoperationen, weil aus Gründen einer möglichst hohen Performance während dieser Operationen z.B. auf das exakte Nachtragen von Indexinformationen, eine Datenverdichtung in Verbindung mit sofortiger Speicherfreigabe oder das Einhalten von Sortierreihenfolgen verzichtet wird. Dabei entstehen auch Auslagerungszeiger, und es kommt zu einer Fragmentierung des Speicherplatzes. Das konkrete Erscheinungsbild der Degenerierungen unterscheidet sich leicht bei unterschiedlichen DBMS.

## 5 WIE - Methoden zur Datenbankreorganisation

Eine Datenbankreorganisation kann explizit veranlaßt werden oder implizit erfolgen. Für die explizit veranlaßte Datenbankreorganisation kann hier nochmals zwischen einer datenbanksystembasierten Reorganisation und einer datenbanksystemintegrierten Reorganisation unterschieden werden.

Das folgende Bild zeigt die Unterteilung der Reorganisationsmethoden für Datenbanken.

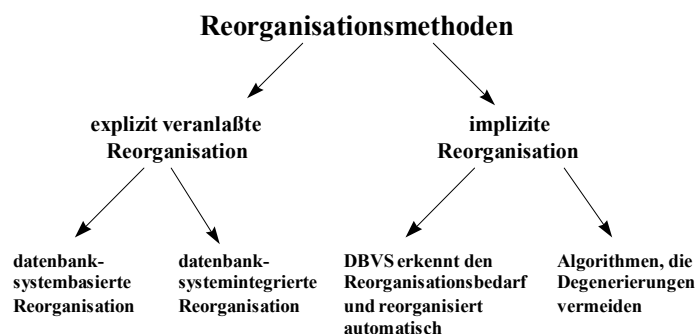


Abb. 2: Einteilung der Reorganisationsmethoden für Datenbanken

### ➤ Explizite Datenbankreorganisation

Zur Durchführung einer manuellen (explizit veranlaßten) Reorganisation bieten Datenbank-Management-Systeme sowohl Möglichkeiten zur datenbanksystembasierten Reorganisation als auch (seltener) solche zur datenbanksystemintegrierten Reorganisation an.

Als **datenbanksystembasierte Reorganisation** sollen Reorganisationsmethoden bezeichnet werden, die in der Regel unter Verwendung von Hilfsprogrammen (Utilities) arbeiten und die aufgesetzt auf einem DBMS realisiert sind (z.B. Dienstprogramme zum Export bzw. Import von Daten verwenden). Die meisten DBMS bieten Möglichkeiten zur datenbanksystembasierten Reorganisation.

Bei **datenbanksystemintegrierten Reorganisationsmethoden** bietet das DBMS (z.B. auf erweiterter SQL-Ebene) Möglichkeiten an, einen Datenbestand oder die entsprechenden Zugriffspfade ohne Aus- und Einlagern des Datenbestands, sozusagen in „einem Schritt“, zu reorganisieren. Die Datenbankreorganisation muß aber vom Benutzer bzw. dem Datenbankadministrator veranlaßt werden. Das DBMS IBM-DB2 bietet bspw. Möglichkeiten zur datenbanksystemintegrierten Reorganisation an.

### ➤ Implizite Datenbankreorganisation

Als **implizit** soll eine Reorganisationsform bezeichnet werden, bei der, im Unterschied zu den beiden oben genannten Methoden, das DBMS selbst oder ein spezieller Reorganisationsprozess (vgl. auch [Soe81]) ein Reorganisationserfordernis erkennt und automatisch eine Reorganisation ausführt. Allerdings verursacht eine solche automatische Datenbankreorganisation i.d.R. Systemlast, welche die Performance, auch zu Zeiten hoher Belastung durch die Benutzer, negativ beeinflussen kann. Deshalb werden derartige Mechanismen in der Praxis oft ungern gesehen. Als eine weitere Form einer impliziten Datenbankreorganisation kann die Nutzung von Algorithmen angesehen werden, die während Änderungsoperationen auf den Speicherstrukturen dafür sorgen, daß diese nicht degenerieren. In diesem Zusammenhang ist teils auch etwas „mutig“ von **Reorganisationsfreiheit** die Rede (siehe auch [SAG97] bzw. [Nuß97]).

## 6 Zusammenfassung und Ausblick

Relationale Datenbank-Management-Systeme (DBMS) zeichnen sich u.a. dadurch aus, daß der Benutzer wenig von Systeminterna, vor allem hinsichtlich der Art und Details der physischen Datenablage, wissen muß und dadurch ein hoher Grad an Datenunabhängigkeit erreicht wird. Trotzdem spielt die physische Datenablage auch hier weiter eine wichtige Rolle. Mit der Problematik: Warum, wann, was, wo und wie reorganisiert werden soll, sind Datenbankadministratoren seit Jahren konfrontiert. Allerdings tritt in jüngster Zeit durch immer größer werdende Datenbanken auf der einen Seite und immer höhere Verfügbarkeitsanforderungen auf der anderen Seite eine Verschärfung der Problematik auf.

Gegenstand laufender Untersuchungen sind u.a. Ansätze zur Reorganisationsbedarfsanalyse, die es ermöglichen soll, Datenbankreorganisationen gezielt auf die degenerierten Bereiche einer Datenbank anzusetzen und um den Nutzen einer Reorganisation abschätzen zu können. Hierzu werden Kennziffern über den Zustand der Strukturen benötigt, die vom DBMS bereitgestellt werden sollten und die dem Datenbankadministrator eine Beurteilung - und Lokalisierung - von Reorganisationserfordernissen gestatten.

### Literaturverzeichnis

- [HR99] Th. Härder, E. Rahm. *Datenbanksysteme: Konzepte und Techniken der Implementierung*. Springer-Verlag, Berlin, Heidelberg, New York, 1999
- [LS93] P. C. Lockemann, J. W. Schmidt. *Datenbankhandbuch*. Springer-Verlag, Berlin, Heidelberg, New York, London, Paris, Tokyo, 1993
- [MHR96] H. Mucksch, J. Holthuis, M. Reiser. Das Data-Warehouse-Konzept - ein Überblick. In *Wirtschaftsinformatik Heft 4/1996*, Verlag Vieweg, Wiesbaden, Juli 1996
- [Nuß97] R. Nußdorfer. Reorganisationsfreiheit in Datenbanksystemen: Bewertung ist anwendungsabhängig. In *Datenbank Fokus, Januar 1997*, IT Verlag für innovative Technologien GmbH, Höhenkirchen, 1997
- [SAG97] *ADABAS D Schattenspeicherkonzept - Reorganisationsfreie Datenhaltung ohne I/O-Engpässe*. SQL GmbH - Software AG, 1997
- [Soe81] L. Soederlund. Concurrent database reorganization – assessment of a powerful technique through modelling. In *Proceedings of the 7<sup>th</sup> Int. Conference on Very Large Data Bases (VLDB)*. Cannes, 1981