

---

# Vorwort

Verteilte und parallele Verarbeitung in Datenbanken erlebte eine erste Blüte in den 80er-Jahren, als parallele Verarbeitung und geografisch verteilte Speicherung von relationalen Daten erstmals intensiv untersucht wurden. In dieser Ära erschienen ein ganze Reihe von fundamentalen Veröffentlichungen und Lehrbüchern, die die Grundlagen für derartige Verfahren legten. In den darauffolgenden Jahren wurden diese Verfahren genutzt, ohne dass dabei ein erneuter Hype zu beobachten war. Dies änderte sich in den letzten Jahren mit Konzepten wie der Cloud-Speicherung und der Analyse in Big-Data-Szenarien, die nun moderne Anforderungen an die Skalierbarkeit mit 20 Jahre altem Lehrbuchwissen konfrontierten. Auch wenn viele der damaligen Entwicklungen auch nach dieser Zeit unverändert gültig sind, fügen diese neuen Szenarien neue Aspekte hinzu, die damals nicht vorhersehbar waren. Es ist nunmehr Zeit für eine neue Generation von Lehrwissen, das die grundlegenden Methoden mit neuen Entwicklungen in gemeinsamem Kontext behandelt. Das vorliegende Buch ist ein Versuch, diesem Bedarf gerecht zu werden.

Die Entstehung dieses Buchs basiert auf zwei Entwicklungen. Einer der Autoren, Erhard Rahm, veröffentlichte bereits 1994 ein umfangreiches Lehrbuch zu Mehrrechnerdatenbanken, das bereits lange aus dem Vertrieb genommen wurde und in den Jahren, als das Thema kein Hype-Thema war, auch nicht aktualisiert wurde. Die anderen beiden Autoren hatten in der letzten Überarbeitung des Buchs „Datenbanken-Implementierungstechniken“ aus Platzmangel die Kapitel zu Aspekten verteilter Datenhaltung, die in den ersten Auflagen noch enthalten waren, stark zusammenkürzen müssen. Gemeinsam haben wir beschlossen, aus dieser Situation das Beste zu machen und Ideen und teilweise auch Inhalte, die nicht mehr als aktueller Lehrbuchinhalt verfügbar waren, gemeinsam in ein neues Lehrbuch zu integrieren – die Idee zum vorliegenden Buch war geboren.

Das vorliegende Buch ist in vier Teile gegliedert. Teil I ist eine breit angelegte Einführung, die die wesentlichen Architekturen, insbesondere Verteilte und Parallele DBS, aber auch neuere NoSQL- und Big-Data-Ansätze, überblicksartig einführt und klassifiziert. Der Teil II vertieft die Methoden der Datenverteilung als Grundlage verteilter Datenzugriffe. Ausführlich wird die Datenallokation für Parallele Datenbanksysteme, inklusive NoSQL-Systemen, betrachtet.

Teil **III** behandelt die Methoden der verteilten und parallelen Anfragebearbeitung. Der Anfragebearbeitung in verteilten Systemen, in Parallelen DBS sowie der parallelen Analyse mit MapReduce ist jeweils ein umfangreiches Kapitel gewidmet.

Der abschließende Teil **IV** ist den Fragen der Konsistenz gewidmet. Verteilte Transaktionen und deren Synchronisation werden in zwei Kapiteln behandelt. Replikation ist zentral für eine effiziente Anfragebearbeitung in verteilten Systemen und zwingend für eine hohe Verfügbarkeit und wird daher in einem Kapitel behandelt. Die speziellen Anforderungen an die Transaktionsverarbeitung für Shared-Disk-Architekturen werden in einem separaten Kapitel ausführlich betrachtet. Die Frage der Konsistenz in Cloud-Datenbanken, die zugunsten der Verfügbarkeit Kompromisse in der Konsistenz des Datenbestandes eingehen, wird in einem gesonderten Kapitel behandelt. Alle Kapitel sind mit Übungsaufgaben zur Vertiefung und Anwendung des erworbenen Wissens versehen.

Gedacht ist das Buch als Basis für eine Master-Vorlesung bzw. fortgeschrittenen Bachelor-Veranstaltung in Informatikstudiengängen. Einzelthemen der verteilten und parallelen Datenhaltung spielen auch eine Rolle in Vorlesungen zur Datenbankimplementierung und zur skalierbaren Analyse im Big-Data-Umfeld. Das Buch ist aber auch für Praktiker geeignet, die sich einen Überblick über Methoden und Techniken der skalierbaren Datenhaltung unter Nutzung von Verteilung und Parallelisierung verschaffen wollen.

Eine ganze Reihe von Mitarbeitern in den drei Arbeitsgruppen haben Kapitel Korrektur gelesen. Unser Dank gilt hier Patrick Arnold, Fabian Benduhn, David Broneske, Victor Christen, Anika Groß, Martin Junghanns, Veit Köppen, Markus Nentwig, Eric Peukert, Ziad Sehili und Christian Wartner. In der Anfangszeit war David Broneske intensiv in das Buchprojekt eingebunden. Sebastian Schultz half uns später beim Setzen der Abbildungen.

Leipzig, Magdeburg und Ilmenau, im Januar 2015

Erhard Rahm, Gunter Saake und Kai-Uwe Sattler

Verteiltes und Paralleles Datenmanagement  
Von verteilten Datenbanken zu Big Data und Cloud  
Rahm, E.; Saake, G.; Sattler, K.-U.  
2015, XIII, 379 S. 149 Abb., Softcover  
ISBN: 978-3-642-45241-3