A Course in In-Memory Data Management

Hasso Plattner

A Course in In-Memory Data Management

The Inner Mechanics of In-Memory Databases

Springer

Hasso Plattner
Hasso Plattner Institute
Potsdam, Brandenburg
Germany

ISBN 978-3-642-36523-2 ISBN 978-3-642-36524-9 (eBook)
DOI 10.1007/978-3-642-36524-9
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013932332

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Why We Wrote This Book

Our research group at the HPI has conducted research in the area of in-memory data management for enterprise applications since 2006. The ideas and concepts of a dictionary-encoded column-oriented in-memory database gained much traction due to the success of SAP HANA as the cutting-edge industry product and from followers trying to catch up. As this topic reached a broader audience, we felt the need for proper education in this area. This is of utmost importance as students and developers have to understand the underlying concepts and technology in order to make use of it.

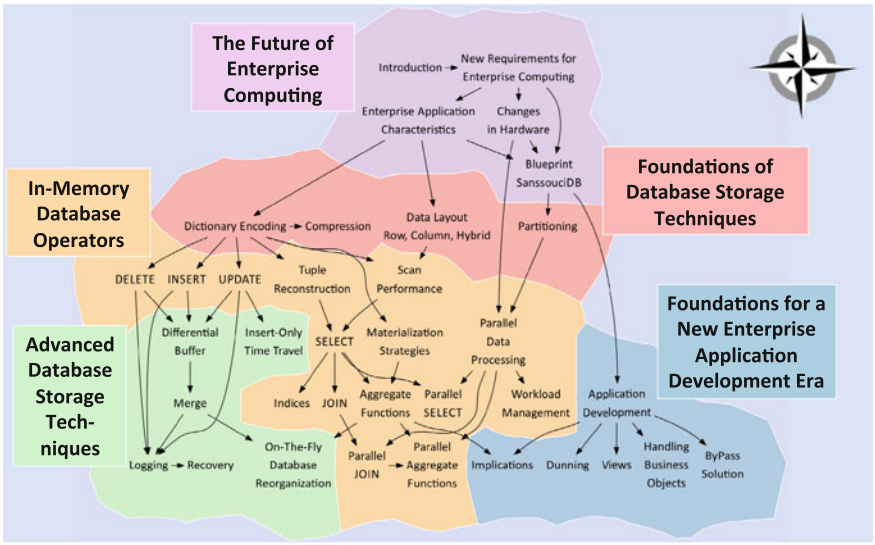
At our institute, we have been teaching in-memory data management in a Master's course since 2009. When I learned about the current movement towards the direction of Massive Open Online Courses, I immediately decided that we should offer our course about in-memory data management to the public. On September 3, 2012 we started our online education with the new online platform http://www.openHPI.de. We granted 2,137 graded certificates to the 13,126 participating learners of the first iteration of the online course. Please feel free to register at openHPI.de to be informed about upcoming lectures.

Several thousand people have already used our material in order to study for the homework assignments and final exam of our online course. This book is based on the reading material that we provided to the online community. In addition to that, we incorporated many suggestions for improvement as well as self-test questions and explanations. As a result, we provide you with a textbook teaching you the inner mechanics of a dictionary-encoded column-oriented in-memory database.

Navigating the Chapters

When giving a lecture, content is typically taught in a one-dimensional sequence. You have the advantage that you can read the book according to your interests. To this end, we provide a learning map, which also reappears in the introduction to

make sure that all readers notice it. The learning map shows all chapters of this book, also referred to as learning units, and shows which topics are prerequisites for which other topics. For example, the learning unit "Differential Buffer" (Chap. 25) is referred to relatively late in the book. Nevertheless, you might already read it earlier. The prerequisites are that you understood the concepts of how "DELETEs", "INSERTs", and "UPDATEs" are conducted without a differential buffer.



The last section of each chapter contains self-test questions. You also find the questions including the solutions and explanations in Sect. 34.3.

The Development Process of the Book

I want to thank the team of our research chair "Enterprise Platform and Integration Concepts" at the Hasso Plattner Institute at the University of Potsdam in Germany. This book would not exist without this team.

Special thanks go to our online lecture core team consisting of *Ralf Teusner*, *Martin Grund*, *Anja Bog*, *Jens Krüger*, and *Jürgen Müller*.

During the preparation of the online lecture as well as during the online lecture itself, the whole research group took care that no email remained unanswered and all reported bugs in the learning material were fixed. Thus, I want to thank the research assistants *Martin Faust*, *Franziska Häger*, *Thomas Kowark*, *Martin Lorenz*, *Stephan Müller*, *Jan Schaffner*, *Matthieu Schapranow*, *David Schwalb*,

Christian Schwarz, *Christian Tinnefeld*, *Arian Treffer*, *Johannes Wust*, as well as our team assistant *Andrea Lange* for their commitment.

During the development process, several HPI bachelor students (Frank Blechschmidt, Maximilian Grundke, Jan Lindemann, Lars Rückert) and HPI master students (Sten Ächtner, Martin Boissier, Ekaterina Gavrilova, Martin Köppelmann, Paul Möller, Michael Wolowyk) supported us during the online lecture preparations. Special thanks go to *Martin Boissier*, *Maximilian Grundke*, *Jan Lindemann*, and *Jasper Schulz*, who worked on all the corrections and adjustments that have to be made when teaching material is enhanced in order to print a book.

Help Improving This Book

We are continuously seeking to improve the learning material provided in this book. If you identify any flaws, please do not hesitate to contact me at hasso.plattner@hpi.uni-potsdam.de.

So far, we received bug reports that resulted in improvements in the learning material from the following attentive readers: Shakir Ahmed, Heiko Betzler, Christoph Birkenhauer, Jonas Bränzel, Dmitry Bondarenko, Christian Butzlaff, Peter Dell, Michael Dietz, Michael Max Eibl, Roman Ganopolskyi, Christoph Gilde, Hermann Grahm, Jan Grasshoff, Oliver Hahn, Ralf Hubert, Katja Huschle, Jens C. Ittel, Alfred Jockisch, Ashutosh Jog, Gerold Kasemir, Alexander Kirov, Jennifer Köenig, Stephan Lange, Francois-David Lessard, Verena Lommatsch, Clemens Müller, Hendrik Müller, Debanshu Mukherjee, Holger Pallak, Jelena Perfiljeva, Dieter Rieblinger, Sonja Ritter, Veronika Rodionova, Viacheslav Rodionov, Yannick Rödl, Oliver Roser, Alice-Rosalind Schell, Wolfgang Schill, Leo Schneider, Jürgen Seitz, David Siegel, Markus Steiner, Reinhold Thurner, Florian Tönjes, Wolfgang Weinmann, Bert Wunderlich, and Dieter Zürn.

We are thankful for any kind of feedback and hope that the learning material will be further improved by the in-memory database community.

Hasso Plattner

Contents

Abbreviations

ATP	Available-to-Promise
BI	Business Intelligence
ccNUMA	Cache-Coherent Non-Uniform Memory Architecture
CPU	Central Processing Unit
DML	Data Manipulation Language
DPL	Data Prefetch Logic
DRAM	Dynamic Random Access Memory
DW	Data Warehouse
e.g.	For example
EPIC	Enterprise Platform and Integration Concepts
ERP	Enterprise Resource Planning
et al.	And others
etc.	Et cetera
ETL	Extract Transform Load
FSB	Front Side Bus
HDD	Hard Disk Drive
HPI	Hasso-Plattner-Institut
HR	Human resources
i.e.	That is
IMC	Integrated Memory Controller
IMDB	In-Memory Database
IO	Index Offsets
IP	Index Positions
MDX	Multidimensional Expression
MIPS	Million Instructions Per Second
NUMA	Non-Uniform Memory Architecture
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
ORM	Object-Relational Mapping
PADD	Parallel Add
PDA	Personal Digital Assistant
QPI	Quick Path Interconnect
RAM	Random Access Memory

RISC	Reduced Instruction Set Computing
SADD	Scalar Add
SIMD	Single Instruction Multiple Data
SRAM	Static Random Access Memory
SSE	Streaming SIMD Extensions
TLB	Translation Lookaside Buffer
UMA	Uniform Memory Architecture

Figures

Tables