Hasso Plattner · Alexander Zeier

In-Memory Data Management

An Inflection Point for Enterprise Applications

Springer

In-Memory Data Management

Hasso Plattner • Alexander Zeier

In-Memory Data Management

An Inflection Point for Enterprise Applications

Springer

Hasso Plattner
Alexander Zeier
Hasso Plattner Institute
Enterprise Platform and Integration Concepts
August-Bebel-Str. 88
14482 Potsdam
Germany
hasso.plattner@hpi.uni-potsdam.de
alexander.zeier@hpi.uni-potsdam.de

ISBN 978-3-642-19362-0 e-ISBN 978-3-642-19363-7
DOI 10.1007/978-3-642-19363-7
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011923529

Cover design: WMX Design GmbH, Heidelberg, Germany

Printed on acid-free paper

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

In Praise of *In-Memory Data Management: An Inflection Point for Enterprise Applications*

Academia

Prof. Christoph Meinel (Hasso Plattner Institute (HPI), Potsdam, Germany)

I'm proud that HPI and the cooperation between HPI and SAP has provided such an inspirational research environment that enabled the young research team around Hasso Plattner and Alexander Zeier to generate valuable and new scientific insights into the complex world of enterprise computations. Even more than that, they developed groundbreaking innovations that will open the door to a new age, the age in which managers can base their decisions on complex computational real-time analysis of business data, and thus will change the way how businesses are being operated.

Prof. David Simchi-Levi (Massachusetts Institute of Technology, Cambridge, USA)

This book describes a revolutionary database technology and many implementation examples for business intelligence and operations. Of particular interest to me are the opportunities opening up in supply chain management, where the need to balance the speed of planning algorithms with data granularity has been a long time obstacle to performance and usability.

Prof. Donald Kossmann (ETH Zurich, Switzerland)

This is the first book on in-memory database systems and how this technology can change the whole industry. The book describes how to build in-memory databases: what is different, what stays the same. Furthermore, the book describes how in-memory databases can become the single source of truth for a business.

Prof. Hector Garcia-Molina (Stanford University, California, USA)

Memory resident data can very significantly improve the performance of data intensive applications. This book presents an excellent overview of the issues and challenges related to in-memory data, and is highly recommended for anyone wishing to learn about this important area.

Prof. Hubert Oesterle (University of St. Gallen, Switzerland)

Technological innovations have again and again been enablers and drivers of innovative business solutions. As database management systems in the 1970s provided the grounds for ERP systems, which then enabled companies in almost all industries to redesign their business processes, upcoming in-memory databases will improve existing ERP-based business solutions (esp. in analytic processing) and will even lead to business processes and services being redesigned again. Plattner and Zeier describe the technical concepts of column- and row-based databases and encourage the reader to make use of the new technology in order to accomplish business innovation.

Prof. Michael Franklin (University of California, Berkeley, USA)

Hardware technology has evolved rapidly over the past decades, but database system architectures have not kept pace. At the same time, competition is forcing organizations to become more and more data-driven. These developments have driven a re-evaluation

of fundamental data management techniques and tradeoffs, leading to innovations that can exploit large memories, parallelism, and a deeper understanding of data management requirements. This book explains the powerful and important changes that are brought about by in-memory data processing. Furthermore, the unique combination of business and technological insights that the authors bring to bear provide lessons that extend beyond any particular technology, serving as a guidebook for innovation in this and future Information Technology revolutions.

Prof. Sam Madden (Massachusetts Institute of Technology, Cambridge, USA)

Plattner and Zeier's book is a thorough accounting of the need for, and the design of, main memory database systems. By analyzing technology trends, they make a compelling case for the coming dominance of main-memory in database systems. They go on to identify a series of key design elements that main memory database system should have, including a column-oriented design, support for multi-core processor parallelism, and data compression. They also highlight several important requirements imposed by modern business processes, including heavy use of stored procedures and accounting requirements that drive a need for no-overwrite storage. This is the first book of it's kind, and it provides a complete reference for students and database designers alike.

Prof. Terry Winograd (Stanford University, California, USA)

There are moments in the development of computer technology when the ongoing evolution of devices changes the tradeoffs to allow a tectonic shift – a radical change in the way we interact with computers. The personal computer, the Web, and the smart phone are all examples where long-term trends reached a tipping point allowing explosive change and growth. Plattner and Zeier present a vision of how this kind of radical shift is coming to enterprise data management. From Plattner's many years of executive experience and development of data management systems, he is able to see the new space of opportunities for users – the potential for a new kind of software to provide managers with a powerful new tool for gaining insight into the workings of an enterprise. Just as the web and the modern search engine changed our idea of how, why, and when we "retrieve information," large in-memory databases will change our idea of how to organize and use operational data of every kind in every enterprise. In this visionary and valuable book, Plattner and Zeier lay out the path for the future of business.

Prof. Warren B. Powell (Princeton University, Princeton, New Jersey, USA)

In this remarkable book, Plattner and Zeier propose a paradigm shift in memory management for modern information systems. While this offers immediate benefits for the storage and retrieval of images, transaction histories and detailed snapshots of people, equipment and products, it is perhaps even more exciting to think of the opportunities that this technology will create for the future. Imagine the fluid graphical display of spatially distributed, dynamic information. Or the ability to move past the flat summaries of inventories of equipment and customer requests to capture the subtle context that communicates urgency and capability. Even more dramatic, we can envision the real-time optimization of business processes working interactively with domain experts, giving us the information-age equivalent of the robots that make our cars and computers in the physical world today.

Prof. Wolfgang Lehner (Technical University of Dresden, Germany)

This book shows in an extraordinary way how technology can drive new applications – a fascinating journey from the core characteristics of business applications to topics of leading-edge main-memory database technology.

Industry

Bill McDermott (Co-CEO, SAP, Newtown Square, Pennsylvania, USA)

We are witnessing the dawn of a new era in enterprise business computing, defined by the near instantaneous availability of critical information that will drive faster decision making, new levels of business agility, and incredible personal productivity for business users. With the advent of in-memory technology, the promise of real-time computing is now reality, creating a new inflection point in the role IT plays in driving sustainable business value. In their review of in-memory technology, Hasso Plattner and Alexander Zeier articulate how in-memory technology can drive down costs, accelerate business, help companies reap additional value out of their existing IT investments, and open the door to new possibilities in how business applications can be consumed. This work is a "must read" for anyone who leverages IT innovation for competitive advantage.

Falk F. Strascheg (Founder and General Partner, EXTOREL, Munich, Germany)

Since the advent of the Internet we have been witnessing new technologies coming up quickly and frequently. It is however rare that these technologies become innovations in the sense that there are big enough market opportunities. Hasso Plattner has proven his ability to match business needs with technical solutions more than once, and this time he presents the perhaps most significant innovation he has ever been working on: Real-Time Business powered by In-Memory Computing. As the ability for innovation has always been one of the core factors for competitiveness this is a highly advisable piece of reading for all those who aim to be at the cutting edge.

Gerhard Oswald (COO, SAP, Walldorf, Germany)

In my role as COO of SAP it is extremely important to react quickly to events and to have instant access to the current state of the business. At SAP, we have already moved a couple of processes to the new in-memory technology described in the book by Hasso Plattner and Alexander Zeier. I'm very excited about the recently achieved improvements utilizing the concepts described in this book. For example, I monitor our customer support messaging system every day using in-memory technology to make sure that we provide our customers with the timely responses they deserve. I like that this book provides an outlook of how companies can smoothly adopt the new database technology. This transition concept, called the bypass solution, gives our existing customer base the opportunity to benefit from this fascinating technology, even for older releases of SAP software.

Hermann-Josef Lamberti (COO, Deutsche Bank, Frankfurt, Germany)

Deutsche Bank has run a prototype with an early versions of the in-memory technology described in the book by Hasso Plattner and Alexander Zeier. In particular, we were able to speed up the data analysis process to detect cross-selling opportunities in our customer database, from previously 45 minutes to 5 seconds. In-memory is a powerful new dimension of applied compute power.

Jim Watson (Managing General Partner, CMEA Capital, San Francisco, California, USA)

During the last 50 years, every IT era has brought us a major substantial advancement, ranging from mainframe computers to cloud infrastructures and smart-phones. In certain decades the strategic importance of one technology versus the other is dramatically different and it may fundamentally change the way in which people do business. This is what a Venture Capitalist has to bear in mind when identifying new trends that are along for the long haul. In their book, Hasso and Alex do not only describe a market-driven innovation from Germany, that has the

potential to change the enterprise software market as a whole, but they also present a working prototype.

Martin Petry (CIO, Hilti, Schaan, Liechtenstein)

Hilti is a very early adopter of the in-memory technology described in the book by Hasso Plattner and Alexander Zeier. Together with SAP, we have worked on developing prototypical new applications using in-memory technology. By merging the transactional world with the analytical world these applications will allow us to gain real-time insight into our operations and allow us to use this insight in our interaction with customers. The benefit for Hilti applying SAP's in-memory technology is not only seen in a dramatic improvement of reporting execution speed - for example, we were able to speed up a reporting batch job from 3 hours to seconds – but even more in the opportunity to bring the way we work with information and ultimately how we service our customers on a new level.

Prof. Norbert Walter (former Chief Economist of Deutsche Bank, Frankfurt, Germany)

Imagine you feel hungry. But instead of just opening the fridge (imagine you don't have one) to get hold of, say, some butter and cheese, you would have to leave the house for the nearest dairy farm. Each time you feel hungry. This is what we do today with most company data: We keep them far away from where we process them. In their highly accessible book, Hasso Plattner and Alexander Zeier show how in-memory technology moves data where they belong, promising massive productivity gains for the modern firm. Decision makers, get up to speed!

Paul Polman (CEO, Unilever, London, UK)

There are big opportunities right across our value chain to use real time information more imaginatively. Deeper, real time insight into consumer and shopper behavior will allow us to work even more closely and effectively with our customers, meeting the needs of today's consumers. It will also transform the way in which we serve our customers and consumers and the speed with which we do it. I am therefore very excited about the potential that the in-memory database technology offers to my business.

Tom Greene (CIO, Colgate-Palmolive Company, New York City, USA)

In their book, Hasso Plattner and Alexander Zeier do not only describe the technical foundations of the new data processing capabilities coming from in-memory, but they also provide examples for new applications that can now be built on top. For a company like Colgate-Palmolive, these new applications are of strategic importance, as they allow for new ways of analyzing our transactional data in real time, which can give us a competitive advantage.

Dr. Vishal Sikka (CTO, Executive Board Member, SAP, Palo Alto, California, USA)

Hasso Plattner is not only an amazing entrepreneur, he is an incredible teacher. His work and his teaching have inspired two generations of students, leaders, professionals and entrepreneurs. Over the last five years, we have been on a fantastic journey with him, from his early ideas on rethinking our core financials applications, to conceiving and implementing a completely new data management foundation for all our SAP products. This book by Hasso and Alexander, captures these experiences and I encourage everyone in enterprise IT to read this book and take advantage of these learnings, just as I have endeavored to embody these in our products at SAP.

To Annabelle and my family

AZ

Foreword

By

Prof. John L. Hennessy (Stanford University, California, USA) and

Prof. David A. Patterson (University of California at Berkeley, USA)

Is anyone else in the world both as well-qualified as Hasso Plattner to make a strong business case for real-time data analytics *and* describe the technical details for a solution based on insights in database design for Enterprise Resource Planning that leverage recent hardware technology trends?

The P of SAP has been both the CEO of a major corporation and a Professor of Computer Science at a leading research institute, where he and his colleagues built a working prototype of a main memory database for ERP, proving once again that Hasso Plattner is a person who puts his full force into the things he believes in. Taking advantage of rapid increases in DRAM capacity and in the number of the processors per chip, SanssouciDB demonstrates that the traditional split of separate systems for Online Transaction Processing (OLTP) and for Online Analytical Processing (OLAP) is no longer necessary for ERP.

Business leaders now can ask ad hoc questions of the production transaction database and get the answer back in seconds. With the traditional divided OLTP/OLAP systems, it can take a week to write the query and receive the answer. In addition to showing how software can use concepts from shared nothing databases to scale across blade servers and use concepts from shared everything databases to take advantage of the large memory and many processors inside a single blade, this book touches on the role of Cloud Computing to achieve a single system for transactions and analytics.

Equally as important as the technical achievement, the "Bill Gates of Germany" shows how businesses can integrate this newfound ability to improve the efficiency and profitability of business, and in a time when so many businesses are struggling to deal with the global problem of markets and supply chains, this instant analytical ability could not be more important. Moreover, if this ability is embraced and widely used, perhaps business leaders can quickly and finely adjust enterprise resources to meet rapidly varying demands so that the next economic downturn will not be as devastating to the world's economy as the last one.

Preface

We wrote this book because we think that the use of in-memory technology marks an inflection point for enterprise applications. The capacity per dollar and the availability of main memory has increased markedly in the last few years. This has led to a rethinking of how mass data should be stored. Instead of using mechanical disk drives it is now possible to store the primary data copy of a database in silicon-based main memory resulting in an orders-of-magnitude improvement in performance and allowing completely new applications to be developed. This change in the way data is stored is having, and will continue to have a significant impact on enterprise applications and ultimately on the way businesses are run. Having real-time information available at the speed of thought provides decision makers in an organization with insights that have, until now, not existed.

This book serves the interests of specific reader groups. Generally, the book is intended for anyone who wishes to find out how this fundamental shift in the way data is managed is affecting, and will continue to affect enterprise applications. In particular, we hope that university students, IT professionals and IT managers, as well as senior management, who wish to create new business processes by leveraging in-memory computing, will find this book inspiring.

The book is divided into three parts:

- *Part I* gives an overview of our vision of how in-memory technology will change enterprise applications. This part will be of interest to all readers.
- *Part II* provides a more in-depth description of how we intend to realize our vision, and addresses students and developers, who want a deeper technical understanding of in-memory data management.
- *Part III* describes the resulting implications on the development and capabilities of enterprise applications, and is suited for technical as well as business-oriented readers.

Writing a book always involves more people than just the authors. We would like to thank the members of our Enterprise Platform and Integration Concepts group at the Hasso Plattner Institute at the University of Potsdam in Germany. Anja Bog, Martin Grund, Jens Krüger, Stephan Müller, Jan Schaffner, and Christian Tinnefeld are part of the HANA research group and their work over the last five years in the field of in-memory applications is the foundation for our book. Vadym Borovskiy, Thomas Kowark, Ralph Kühne, Martin Lorenz, Jürgen Müller, Oleksandr Panchenko, Matthieu Schapranow, Christian Schwarz, Matthias Uflacker, and Johannes Wust

also made significant contributions to the book and our assistant Andrea Lange helped with the necessary coordination.

Additionally, writing this book would not have been possible without the help of many colleagues at SAP. Cafer Tosun in his role as the link between HPI and SAP not only coordinates our partnership with SAP, but also actively provided sections for our book. His team members Andrew McCormick-Smith and Christian Mathis added important text passages to the book. We are grateful for the work of Joos-Hendrik Böse, Enno Folkerts, Sarah Kappes, Christian Münkel, Frank Renkes, Frederik Transier, and other members of his team. We would like to thank Paul Hofmann for his input and for his valuable help managing our research projects with American universities. The results we achieved in our research efforts would also not have been possible without the outstanding help of many other colleagues at SAP. We would particularly like to thank Franz Färber and his team for their feedback and their outstanding contributions to our research results over the past years. Many ideas that we describe throughout the book were originally Franz's, and he is also responsible for their implementation within SAP. We especially want to emphasize his efforts.

Finally, we want to express our gratitude to SAP CTO Vishal Sikka for his sponsorship of our research and his personal involvement in our work. In addition, we are grateful to SAP COO Gerhard Oswald and SAP Co-CEOs Jim Hagemann Snabe and Bill McDermott for their ongoing support of our projects.

We encourage you to visit the official website of this book. The website contains updates about the book, reviews, blog entries about in-memory data management, and examination questions for students. You can access the book's website via:

no-disk.com

Contents

Introduction

Over the last 50 years, advances in Information Technology (IT) have had a significant impact on the success of companies across all industries. The foundations for this success are the interdependencies between business and IT, as they not only address and ease the processing of repetitive tasks, but are the enabler for creating more accurate and complete insights into a company. This aspect has often been described and associated with the term real-time as it suggests that every change that happens within a company is instantly visible through IT.

We think that significant milestones have been reached towards this goal throughout the history of enterprise computing, but we are not there, yet. Currently, most of the data within a company is still distributed throughout a wide range of applications and stored in several disjoint silos. Creating a unified view on this data is a cumbersome and time-consuming procedure. Additionally, analytical reports typically do not run directly on operational data, but on aggregated data from a data warehouse. Operational data is transferred into this data warehouse in batch jobs, which makes flexible, ad-hoc reporting on up-to-date data almost impossible. As a consequence, company leaders have to make decisions based on insufficient information, which is not what the term real-time suggests. We predict this is about to change as hardware architectures have evolved dramatically in the last decade. Multi-core processors and the availability of large amounts of main memory at low cost are creating new breakthroughs in the software industry. It has become possible to store data sets of whole companies entirely in main memory, which offers performance that is orders of magnitudes faster than traditional disk-based systems. Hard disks will become obsolete. The only remaining mechanical device in a world of silicon will soon only be necessary for backing up data. With in-memory computing and insert-only databases using row- and column-oriented storage, transactional and analytical processing can be unified. High performance in-memory computing will change how enterprises work and finally offer the promise of real-time computing.

As summarized in Figure I, the combination of the technologies mentioned above finally enables an iterative link between the instant analysis of data, the prediction of business trends, and the execution of business decisions without delays.

H. Plattner and A. Zeier, *In-Memory Data Management: An Inflection Point for Enterprise Applications*, DOI 10.1007/978-3-642-19363-7_1, © Springer-Verlag Berlin Heidelberg 2011

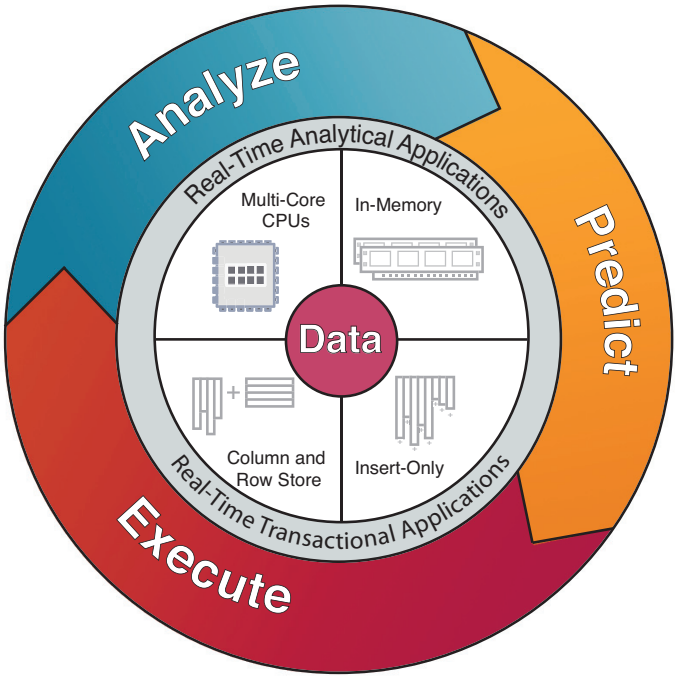

Figure I: Enterprise Performance In-Memory Circle (EPIC)

How can companies take advantage of in-memory applications to improve the efficiency and profitability of their business? We predict that this break-through innovation will lead to fundamentally improved business processes, better decision-making, and new performance standards for enterprise applications across industries and organizational hierarchies. We are convinced that in-memory technology is a catalyst for innovation, and the enabler for a level of information quality that has not been possible until now. In-memory enterprise data management provides the necessary equipment to excel in a future where businesses face ever-growing demands from customers, partners, and shareholders. With billions of users and a hundred times as many sensors and devices on the Internet, the amount of data we are confronted with is growing exponentially. Being able to quickly extract business-relevant information not only provides unique opportunities for businesses; it will be a critical differentiator in future competitive markets.

With in-memory technology, companies can fully leverage massive amounts of data to create strategic advantage. Operational business data can be interactively analyzed and queried without support from the IT department, opening up completely new scenarios and opportunities.

Consider financial accounting, where data needs to be frequently aggregated for reporting on a daily, weekly, monthly, or annual basis. With in-memory data management, the necessary filtering and aggregation can happen in real time. Accounting can be done anytime and in an ad-hoc manner. Financial applications

will not only be significantly faster, they will also be less complex and easier to use. Every user of the system will be able to directly analyze massive amounts of data. New data is available for analysis as soon as it is entered into the operational system. Simulations, forecasts, and what-if scenarios can be done on demand, anytime and anywhere. What took days or weeks in traditional disk-based systems can now happen in the blink of an eye. Users of in-memory enterprise systems will be more productive and responsive.

The concepts presented in this book create new opportunities and improvements across all industries. Below, we present a few examples:

- *Daily Operations*: Gain real-time insight into daily revenue, margin, and labor expenses.
- *Competitive Pricing*: Intuitively explore impact of competition on product prizing to instantly understand impact to profit contribution.
- *Risk Management*: Immediately identify high-risk areas across multiple products and services and run what-if scenario analyses on the fly.
- *Brand and Category Performance*: Evaluate the distribution and revenue performance of brands and product categories by customer, region, and channel at any time.
- *Product Lifecycle and Cost Management*: Get immediate insight into yield performance versus customer demand.
- *Inventory Management*: Optimize inventory and reduce out-of-stocks based on live business events.
- *Financial Asset Management*: Gain a more up-to-date picture of financial markets to manage exposure to currencies, equities, derivatives, and other instruments.
- *Real-Time Warranty and Defect Analysis*: Get live insight into defective products to identify deviation in production processes or handling.

In summary, we foresee in-memory technology triggering the following improvements in the following three interrelated strategic areas:

- *Reduced Total Cost of Ownership*: With our in-memory data management concepts, the required analytical capabilities are directly incorporated into the operational enterprise systems. Dedicated analytical systems are a thing of the past. Enterprise systems will become less complex and easier to maintain, resulting in less hardware maintenance and IT resource requirements.
- *Innovative Applications*: In-memory data management combines high-volume transactions with analytics in the operational system. Planning, forecasting, pricing optimization, and other processes can be dramatically improved and supported with new applications that were not possible before.
- *Better and Faster Decisions*: In-memory enterprise systems allow quick and easy access to information that decision makers need, providing them with new ways to look at the business. Simulation, what-if analyses, and planning can be performed interactively on operational data. Relevant information is instantly accessible and the reliance on IT resources is reduced. Collaboration within and across organizations is simplified and fostered. This can lead to a

much more dynamic management style where problems can be dealt with as they happen.

At the research group Enterprise Platform and Integration Concepts under the supervision of Prof. Dr. Hasso Plattner and Dr. Alexander Zeier at the Hasso Plattner Institute (HPI) we have been working since 2006 on research projects with the goal of revolutionizing enterprise systems and applications. Our vision is that in-memory computing will enable completely new ways of operating a business and fulfill the promise of real-time data processing. This book serves to explain in-memory database technology and how it is an enabler for this vision. We go on to describe how this will change the way enterprise applications are developed and used from now on.

PART I – An Inflection Point for Enterprise Applications

For as long as businesses have existed, decision makers have wanted to know the current state of their company. As businesses grow, however, working out exactly where the money, materials, and people go becomes more and more complicated. Tools are required to help to keep track of everything. Since the 1960s, computers have been used to perform this task and complex software systems called enterprise applications have been created to provide insights into the daily operations of a company. However, increasing data volumes have meant that by the turn of the 21st century, large organizations were no longer always able to access the information they required in a timely manner.

At the heart of any enterprise application is the database management system, responsible for storing the myriad of data generated by the day-to-day operations of a business. In the first part of this book, we provide a brief introduction to enterprise applications and the databases that underlie them. We also introduce the technology that we believe has created an inflection point in the development of these applications. In Chapter 1 we explain the desirability, feasibility, and viability of in-memory data management. Chapter 2 introduces the complexity and common data access patterns of enterprise applications. Chapter 3 closes the first part with the description of SanssouciDB, our prototypical in-memory database management system.

1 Desirability, Feasibility, Viability – The Impact of In-Memory

Abstract Sub-second response time and real-time analytics are key requirements for applications that allow natural human computer interactions. We envision users of enterprise applications to interact with their software tools in such a natural way, just like any Internet user interacts with a web search engine today by refining search results on the fly when the initial results are not satisfying. In this initial chapter, we illustrate this vision of providing business data in real time and discuss it in terms of desirability, feasibility, and viability. We first explain the desire of supplying information in real time and review sub-second response time in the context of enterprise applications. We then discuss the feasibility based on in-memory databases that leverage modern computer hardware and conclude by demonstrating the economic viability of in-memory data management.

In-memory technology is set to revolutionize enterprise applications both in terms of functionality and cost due to a vastly improved performance. This will enable enterprise developers to create completely new applications and allow enterprise users and administrators to think in new ways about how they wish to view and store their data. The performance improvements also mean that costly workarounds, necessary in the past to ensure data could be processed in a timely manner, will no longer be necessary. Chief amongst these is the need for separate operational and analytical systems. In-memory technology will allow analytics to be run on operational data, simplifying both the software and the hardware landscape, leading ultimately to lower overall cost.

1.1 Information in Real Time – Anything, Anytime, Anywhere

Today's web search engines show us the potential of being able to analyze massive amounts of data in real time. Users enter their queries and instantly receive answers. The goal of enterprise applications in this regard is the same, but is barely reached. For example, call center agents or managers are looking for specific pieces of information within all data sources of the company to better decide on products to offer to customers or to plan future strategies. Compared to web search with

H. Plattner and A. Zeier, *In-Memory Data Management: An Inflection Point for Enterprise Applications*, DOI 10.1007/978-3-642-19363-7_2,