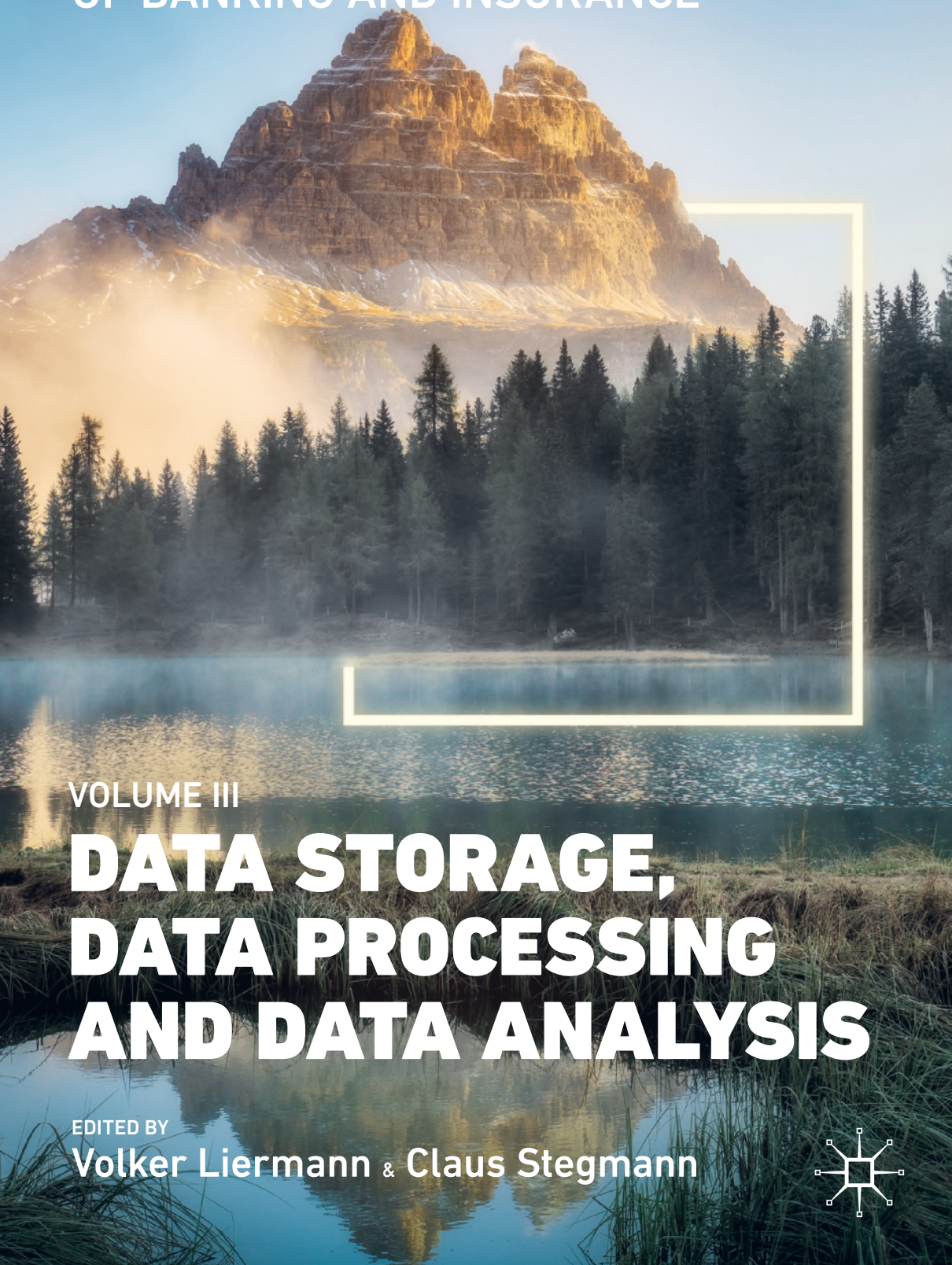
THE DIGITAL JOURNEY
OF BANKING AND INSURANCE



VOLUME III

DATA STORAGE, DATA PROCESSING AND DATA ANALYSIS

EDITED BY
Volker Liermann & Claus Stegmann

The Digital Journey of Banking and Insurance, Volume III

"Virtually all financial institutions have embarked on ambitious digital journeys, both to provide better products and customer experience more efficiently and in response to the threat of industry disruption by FinTech competitors. There is no doubt that there will be winners, and there will be losers. I am convinced that *The Digital Journey of Banking and Insurance* series is indispensable reading for the future winners."

—Thomas C. Wilson, *CEO, President and Country Manager at Allianz Ayudhya*

"*Data Storage, Processing, and Analysis*, the last volume of *The Digital Journey of Banking and Insurance*, gives in-depth insights into technological aspects which is essential for successful digital transformation."

—Dr. Carsten Stolz, *CFO Baloise Group*

"Technological aspects are essential for successful digital transformation and so I like to get in-depth insights by *Data Storage, Processing, and Analysis*, the last volume of *The Digital Journey of Banking and Insurance*."

—Gerhard Lahner, *COO of Vienna Insurance Group*

"We do remember when we started our digital journey, but we do not know when it will be over. Therefore, we are definitely in the middle. The book series *The Digital Journey of Banking and Insurance* is a must-read for of all of us."

—Christian Peter Kromann, *CEO, SimCorp*

"Although the subjects described in this book are technical, the authors find a way to explain them in a comprehensible way. An up-to-date book for this subject."

—Bernhard Hodler, *Former CEO Julius Baer Group*

Volker Liermann · Claus Stegmann
Editors

The Digital Journey of Banking and Insurance, Volume III

Data Storage, Data Processing and Data Analysis

palgrave
macmillan

Editors
Volker Liermann
ifb SE
Grünwald, Germany

Claus Stegmann
ifb Americas, Inc.
Charlotte, NC, USA

ISBN 978-3-030-78820-9 ISBN 978-3-030-78821-6 (eBook)
https://doi.org/10.1007/978-3-030-78821-6

Cover credit: Stutterstock/Blue Planet Studio

This Palgrave Macmillan imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Acknowledgments

The three-book series was the natural next step from the book "The Impact of Digital Transformation and FinTech on the Finance Professional" and an exciting project for us. We look back with gratitude at the many discussions with clients, partners, and colleagues at ifb. Without this vital community, such an undertaking would not be possible.

We would first like to thank all contributors (clients, partners, and colleagues), whose expertise was invaluable in exploring and formulating such a comprehensive work with a wide overview and deep insights. Their insightful feedback helped us to sharpen this work to this amazing level.

In addition, we would like to thank Tula Weis and her team from Palgrave Macmillan for their advice and support in this project.

We like to thank Satzanstalt for supporting us in the development and realization of the book covers idea.

We would also like to thank our colleagues Julia Horstmann, Davin Radermacher, and Jenny Klein for their support in all the small, but important things that make such an undertaking a success.

Cologne, Germany
March 2021

Volker Liermann
Claus Stegmann

Introduction to Volume III—Data Storage, Data Processing and Data Analysis

The business models of financial sector companies were always and are still information-driven. Digital transformation and the stronger customer focus (driven by the fintech companies and Big Tech) demand information on customer behavior (NBO,[1] NBA[2]). These behavioral customer patterns are the key to continuous revenue generation.

John Naisbitt wrote in his well-known book Megatrends, "We are drowning in information but starved for knowledge." (see Naisbitt, 1982). Handling data has become a key—if not the most decisive—capability and skill an organization in the financial sector needs. Handling data (including analyzing data) is the task to transform information into knowledge.

To be more precise, handling data means collecting, storing, and transforming (and analyzing) data.

Three major trends—especially in external digitalization—are driving the data handling process: Trend A: increase in available data, Trend B: accelerated speed in data processing, Trend C: special structures for optimized storage and querying of complex and unnormalized[3] data structures. These three trends are mirrored in technologies: Trend A is reflected in the new cluster databases like Hadoop (or AWS-S3, Google Bigtable), making the handling

[1] Next best offer (see May, 2019).

[2] Next best action.

[3] The process of organizing the fields and tables of a relational database is called database normalization. Normalization helps to minimize redundancy and dependency.

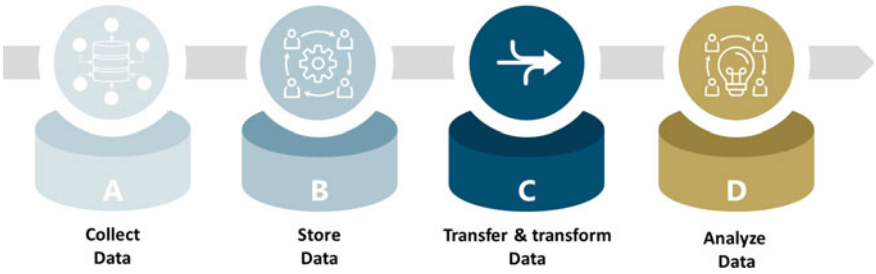


Fig. 1 Data process (© ifb SE)

of high data volumes possible and affordable. Trend B is shown in streaming technology (like Kafka, see Steurer, 2021) making real- or near-time data provision possible. Streaming technology was in place long before the digital transformation sped up, and the first steps with Kafka Standalone faced several challenges. The modern architecture concepts like Lambda, Kappa and Delta architectures (see Krätz and Morawski, Data Infrastructures—Lambda Architecture and Other Architectures, 2021) combine traditional architecture patterns providing stability with the dynamism and speed of streaming technology. Other facets of Trend B are in-memory databases (IMDB), making it feasible to handle huge data volumes in the blink of an eye. Trend C materializes in different specialized databases (like document-based databases and graph databases) as well as in distributed ledgers.

Mass Data and Data Availability

Trend A (increase in available data) has its origin in the earlier days of the internet. The internet was growing, and Google needed to store this growing data as a key capability of a search engine. Google was looking for affordable mass data storage. Google File System (GFS) is a proprietarily distributed file system.[4] Some of its components influenced Hadoop and its derivatives and advancement.

Now these two elements, (A) growing volume of available data and (B) technology to handle this data (at reasonable cost), started to interact and to scale up with ever more continuous acceleration. Google transformed its business model from a search engine (with a focus on advertising) to other

[4] Google File System targets an efficient, reliable access to data stored in large clusters of commodity hardware.

data-driven revenue models (DDRM). Imitating this success, a whole sector of data-driven revenue business models has arisen.

After seeing the success of DDBM, other sectors started to explore their opportunities, transforming data into benefits for their existing business model. The setup is made up of four kinds of sources: (1) existing data (data unused until now), (2) newly collected data (which was lost because it was deleted or not properly archived), (3) external data (e.g., Google, Facebook), (4) the intelligent linkage of the previous three (see Fig. 2).

Fig. 2 Data process (© ifb SE)

The Internet of Things (IoT) is a collective term for technologies of a global infrastructure within information societies. It makes it possible to network physical and virtual objects with each other and to let them work together through information and communication technologies. The Internet of Things produces an immense volume of data every day.

Another aspect coming with growing data availability is the way data is available. The traditional association was structured data (in a relational database, maybe differentiated by dimensions and key figures). The newly collected data was digitally (or electronically) available but in an unstructured way.

Speed and Streaming

Dynamism in modern (streaming/integrating) architecture has two dimensions. Operational dynamism (the data availability has improved) and a change-enabling dynamism (due to the micro-service component, changes (examples can be found in (Steurer, 2021) can be deployed much faster).

When looking at outside digitalization, the need to speed up is more obvious than in inside digitalization. Identifying and solving the right problem for the customer at the right time is the only way to stay near the customer in an increasingly competitive market (if this is the target of the company's business model[5]).

Speed for its own sake does not necessarily provide benefits. For example, instant payment[6] was not well-received in the early days in 2017. In contrast, Zalando (and other e-commerce companies) offer a two-week payment term and bring the payment closer to a debit card payment.

It must be considered that (not only) Millennials[7] and Generation Z[8] use communication and media differently and dispose disparately of availability of goods and services. Young people's changing demands and utilization have always impacted older generations (in this case, Generation X[9] and Baby Boomers[10]).

The important transactions in life, like taking out a mortgage, are more often perceived as special occasions that should be celebrated. They are more likely to be performed in a branch office, involving direct interactions with another human being. In contrast, other services like checking an account (or checking all accounts across banks via an aggregator app) or a consumer credit (which is less of a financial transaction than a surplus to buy goods) do not require direct interaction. In retail banking transactions, the right timing or, more accurately, the right context will win the deal. The enablers here are the relevant data and infrastructure with a speed capability.

Things look different in corporate banking, for now. Corporates are changing, and the finance departments of medium and large companies will

[5] Some analysts see a wider spread coming for institutes providing products (manufacturer) and institutes composing products to a solution (orchestrator).

[6] SEPA Instant Payment.

[7] Or Generation Y: Generation born in the period from the early 1980s to the late 1990s.

[8] Gen Z for short, this is the simplified term used to describe the successor generation to Generation Y.

[9] The term Generation X (also Gen X) usually refers to the generation following the Baby Boomers.

[10] Baby Boomers refers to the generation born during the periods of rising birth rates (the "baby boom") following World War II or wars in other countries.

go through a transformation in the next decade, demanding more standardized and easier to use interfaces (API[11]). These digital improvements will serve the business model targets of corporates and will put speed pressure on the availability of financial transactions.

Trends A & B are relevant for insurance companies in the same way.

These trends will affect and change the infrastructure (driven by outside digitalization). It does not make sense to work with two kinds of infrastructures (in terms of employees' technological skills). It is therefore expected that the modern streaming architecture patterns and infrastructures will become available to the finance department (even without a business case[12]).

To summarize, modern streaming architectures (like Delta architecture) are about to be driven by outside digitalization (customer-related) and by inside digitalization (optimizing processes and data flows).

In-Memory Databases

An in-memory database (IMDB) is an innovative database management system that uses the main memory of a computer as data storage. In-memory databases can be distinguished from conventional database management systems, which use hard disk drives for this purpose. Many standard software vendors[13] and a number of open-source frameworks[14] offer in-memory databases as a tool or toolset. Most of the in-memory databases have also implemented column-oriented storage and query to improve performance.

In-memory databases are fast and extremely valuable in selected applications, but the required hardware can be costly. A function-driven distribution of data among the different types of databases (hot, warm, and cold storage), called data tiering, is extremely useful and can reduce costs.

[11] Application programming interface.

[12] A business case for only speeding up accounting processes is a challenge because the value of timely information will always be viewed differently.

[13] SAP—HANA (for details see Kopic et al., 2019), Oracle—TimesTen, Microsoft—Hekaton.

[14] Apache Ignite, Redis, VoltDB.

Unstructured Data

Addressing the challenge of unstructured data, a new kind of database (NoSQL[15]) was established. Well-known implementations include Riak, Apache Cassandra, CouchDB, MongoDB (see Bialek, 2021), and Redis. Most of the cluster databases like Hadoop can be file-based and therefore have at least components of NoSQL databases.

Unstructured data can be text (like in an email or a PDF file) already containing electronic characters, but it also covers text scans or other kinds of graphics, videos, and spoken recordings. Text scans, graphics, videos and spoken recordings can be classified as raw material. Voice detection, OCR[16] and image processing are well-established tools to transform the raw formats into a character-driven (or object-driven) format. Natural language processing is the standard toolset to transform content from a character-driven format to structured data.

The ability to process unstructured data (especially character-driven data) opens up an endless stream of data that can be transferred into structured information using natural language processing.

The new availability of data generates demand to put the different entities in a context of interconnectedness by showing the connections ("Everything is connected" see also (Enzinger & Grossmann, 2019). To handle this connection, a graph[17] (originated from the mathematical object[18]) is an extremely powerful tool. Another new database class has been established to handle this connection data: graph databases and special-purpose query languages like Cypher (for more details see Bajer et al., 2021).

Distributed Ledger

In terms of storing data, distributed ledger with all its facets is a huge subject. Starting in 2008 with the Nakamotos Bitcoin whitepaper (Nakamoto, 2008), Bitcoin has gained visibility and attention. After the price peak at the end of 2017, it has now (March 2021) moved to new high values. In Bitcoin, the

[15] NoSQL (Not Only SQL) refers to databases that follow a non-relational approach and break with the paradigm of relational databases.

[16] Optical character recognition.

[17] A graph consists of nodes and edges (see Biggs et al., 1986).

[18] Graph Theory is a branch of discrete mathematics and theoretical computer science (see Diestel, 2017).

transfer and storage of values is paramount, or, if we oversimplify it, it is payment without an intermediator.

The next topic in the distributed ledger universe is smart contracts and the enrichment of the storage and payment functions by an almost endless set of functions that can implement the intentions in or behind legal contracts (dividend payments, digital rights, …). The most popular technology in the domain of public blockchains[19] is Ethereum. To implement private blockchains, Corda[20] and Hyperledger[21] are popular frameworks.

The different alignments of distributed ledger are summarized under the umbrella of the abbreviation DeFi (decentralized finance), highlighting the decentralized and distributed approach. Decentralized finance emphasizes the lack of intermediator, making it an experimental, novel form of financial market based on smart contracts and decentralized autonomous organizations.

Smart contracts enable participants to implement tokens with a variety of characteristics and features. The tokenization of real-world assets like Everledger (Foreverhold Ltd., 2020) has become more and more popular, while the mirroring of real-world assets into a distributed ledger can digitalize and optimize processes happening around real-world assets.

Another important aspect in the digital world is self-sovereign identity (SSI). This also allows control over the way personal data is shared and used. Self-sovereign identity allows a person, organization, or machine to create and fully control a digital identity without requiring permission from an intermediary or central entity Distributed ledger technology has certain technological advantages, making the distributed ledger an optimal platform for implementing SSI.

Data Analysis

The last part of the data journey in Fig. 1 is data analysis. In this subject area, data mining and analysis can be applied as well as machine learning and deep learning. There are various baskets of different open-source proprietary tools, frameworks, and platforms available, and they are improving and

[19] The concept of a public blockchain is behind most major cryptocurrencies. Access to this blockchain variant is open to any participant, so anyone who wants to execute transactions, validate blocks, and view the entire history of the blockchain is allowed to do so.

[20] See Corda, 2020.

[21] See The Linux Foundation, 2020.

transforming continuously to serve the purpose of transforming data into information to support the core business model of the company.

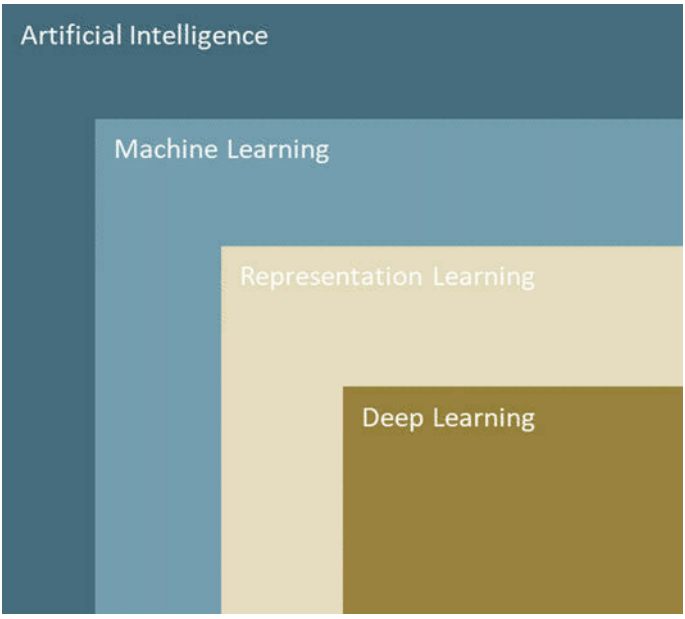


Fig. 3 Overview artificial intelligence (© ifb SE)

Figure 3 shows the traditional subset visualization of artificial intelligence. Machine learning has evolved from different statistical disciplines. A good definition of machine learning is given in (Chakraborty and Joseph, 2017) and an introduction can be found in (Liermann et al., Mathematical Background of Machine Learning, 2019). Representation learning[22] aims to replace manual feature engineering, using techniques allowing feature detection or other raw data classification by representations. Deep learning is a subset of machine learning methods that implements artificial neural networks (NN) by forming an internal structure with numerous hidden layers between the input layer and the output layer. See (Goodfellow, Bengio und Courville, 2014) for a comprehensive introduction to deep learning and (Liermann et al., Deep Learning—an Introduction, 2019) for an introduction to deep learning.

[22] Sometimes also referred to as feature learning.

Overview of Book Series "The Digital Journey of Banking and Insurance"

This book is the third volume of the three-volume book series "The Digital Journey of Banking and Insurance." The first volume "Disruption and DNA" focuses on change and the things staying stable in the banking and insurance market (outside view) as well as the effect on accounting, risk management, and regulatory departments (inside view). The inside view is rounded off with an analysis of cultural alterations.

The second volume "Digitalization and Machine Learning Applications" mainly emphasizes use cases as well as the methods and technologies applied (such as processes, leveraging computational power and machine learning models).

This volume "Data Storage, Processing and Analysis," the last one of the series, considers how to deal with data. The angle shifts over the volumes from a business-driven approach in "Disruption and DNA" to a strong technical focus in "Data Storage, Processing and Analysis," leaving "Digitalization and Machine Learning Applications" in-between with business and technical aspects.

Literature

Bajer, Krystyna, Sascha Steltgens, Anne Seidlitz, and Bastian Wormuth. 2021. "Graph Databases." In *The Digital Journey of Banking and Insurance, Volume III—Data Storage, Processing, and Analysis*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

Bialek, Boris. 2021. "Digitization and MongoDB." In *The Digital Journey of Banking and Insurance, Volume III—Data Storage, Processing, and Analysis*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

Norman L. Biggs, E. Keith Lloyd, and Robin J. Wilson. 1986. *Graph Theory 1736–1936*. London: Oxford University Press.

Chakraborty, Chiranjit, and Andreas Joseph. 2017. *Staff Working Paper No. 674—Machine Learning at Central Banks*. London: Bank of England.

Corda. 2020. *Corda*. Accessed December 15, 2020. https://www.corda.net/.

Diestel, R. (2017). *Graph Theory*. Springer.

Enzinger, Philipp, and Stefan Grossmann. 2019. "Managing Internal and External Network Complexity." In *The Impact of Digital Transformation and Fintech on the Finance Professional*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

Foreverhold Ltd. 2020. *Everledger*. Accessed December 15, 2020. https://www.everledger.io/.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2014. *Deep Learning.* http://www.deeplearningbook.org. MIT Press. http://www.deeplearningbook.org.

Kopic, Eva, Bezu Teschome, Thomas Schneider, Ralph Steurer, and Sascha Florin. 2019. "In-Memory Databases and Their Impact on Our (Future) Organizations." In *The Impact of Digital Transformation and Fintech on the Finance Professiona*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

Krätz, Dennis, and Michael, Morawski. 2021. "Architecture Patterns—Batch & Real Time Capabilities." In *The Digital Journey of Banking and Insurance, Volume III—Data Storage, Processing, and Analysis*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

Liermann, Volker, Sangmeng Li, and Norbert Schaudinnus. 2019. "Deep Learning—an Introduction." In *The Impact of Digital Transformation and Fintech on the Finance Professional*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

Liermann, Volker, Sangmeng Li, and Norbert Schaudinnus. 2019. "Mathematical Background of Machine Learning." In *The Impact of Digital Transformation and Fintech on the Finance Professional*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

May, Uwe. 2019. "The Concept of the Next best Action/Offer in the age of Customer Experience." In *The Impact of Digital Transformation and Fintech on the Finance Professional*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

Naisbitt, J. (1982). *Megatrends: Ten New Directions Transforming Our Lives*. Warner Books.

Nakamoto, Satoshi. 2008. *Bitcoin—a Peer-to-Peer Electronic Cash System.*

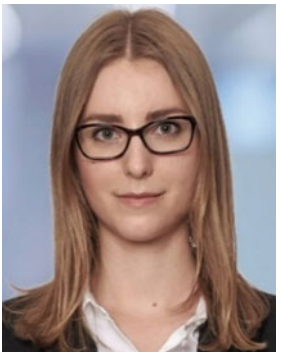
Steurer, Ralph. 2021. "Kafka—Real-Time Streaming for the Finance Industry." In *The Digital Journey of Banking and Insurance, Volume III: Data Storage, Processing, and Analysis*, edited by Volker Liermann and Claus Stegmann. New York: Palgrave Macmillan.

The Linux Foundation. 2020. *Hyperledger.* Accessed December 15, 2020. https://www.hyperledger.org/.

Contents

Notes on Contributors

Eljar Akhgarnush Implementation Consultant at ifb group since April 2018, gained knowledge in software and financial sector topics during his studies as well as the various positions he has since held. Starting out with a focus on financial supervision and regulation, he soon shifted his attention to the technical side and agile project management. After having received his B.Sc. in business administration from CAU Kiel, he gained his M.Sc. in international economics and policy consulting at OvGU in Magdeburg. He has since remained keen on exchanging knowledge and exploring new areas.



Krystyna Bajer has been a Consultant at ifb group in the Data Information Team since the beginning of 2020. She holds a degree in business mathematics and financial services risk management and has been able to gain experience in the banking and insurance industry. Since October 2020, she has been assisting in a project at an insurance company in the area of data architecture.

Boris Bialek Global Head of Enterprise Modernization, leads the industry practices at MongoDB and specifically focuses on the modernization of banking solutions. His work focus is digital transformation and true innovation implementing exciting solutions, be it a new mobile payment platform for a US banking group or risk and treasury platform for a G-SIB that allows real-time data reconciliation. Before joining MongoDB, he worked for many years with FIS, IBM, Dell and Compaq Computers. He was one of the founding members of the SAP LinuxLab, implementing the first ever Linux client with SAP. He obtained an M.Sc. degree from Karlsruhe Institute of Technology.

Fabian Bruse Director, has worked at ifb group since 2011. He started his career in the regulatory reporting sector as a software tester and later moved on to SAP BW and SAP BA development with a particular focus on IRR and CRA modules for customers in Germany and Luxembourg. His more recent projects include modern ETL and reporting processes where he has the role of a—partially remote—Scrum Master (PSM2 certified). Since 2017, he has also coordinated the technical part of the ifb Blockchain Team and administrated the ifb Hyperledger system on Kubernetes. Fabian has a degree in physics from the University of Bonn.

Marie Kristin Czwalina Senior Consultant, has been working at ifb group in the Core Banking Team since 2020. She deals with further development in the area of innovation management and digitalization topics around the areas of sales and processing in banks, and specializes in process automation, customer-oriented advice and the implementation of GDPR requirements in IT applications. She studied business informatics with

a focus on IT management in her master's degree at the FOM University of Applied Sciences.

Milan den Heijer is a Managing Consultant at ifb group. He has been working in IT consulting in the financial sector for six years. His focus is on data modeling, ETL modeling and data warehousing. In these fields, he has gained experience with platforms such as SAP BW, SAP FSDM and SAP HANA. Moreover, he has experience in the fields of master data management, data governance and metadata management. He has a background in physics and astronomy.

Dr. Johannes Floß has been a consultant with ifb group since 2019. He mainly works on IFRS 17 implementation projects with a specialization on SAP FPSL and SAP PaPM. As a second topic, he develops data science tools for the insurance business, e.g. the prediction of churn rates with the help of machine learning algorithms. Before his time at ifb, Johannes was a research assistant at the University of Toronto's Centre for Quantum Information and Quantum Computing, studying fundamental problems like quantum chaos and light-matter interaction. He holds a Ph.D. in chemistry from the Weizmann Institute of Science, Israel.

Jens Freche Managing Consultant and Team Lead Data Management, has worked at ifb group since 2011. He started his career in data integration for Oracle and SAP systems. Later, he moved on to SAP BW development, data governance, SAP FPSL and IFRS 17 standard for customers in various countries including Germany, Luxembourg, Switzerland, the USA, Japan, Chile and Israel. Since 2018, he has been the Head of the Data Lineage and Data Governance Team. Jens has a Master of Science in Mathematics from the University of Applied Science Aachen.

Dennis Kraetz Partner at ifb group, has been working in the consulting industry for more than 15 years, focusing on finance transformations in the financial services sector. Within ifb, he leads the consulting practice for Cross Industries, developing solutions in the context of architecture, information and transformation management. In recent years, he has been concentrating on emerging architecture patterns and their application in the banking and insurance industry. Dennis holds a degree in business administration.

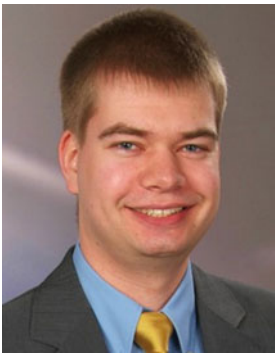
Heinrich Krebs Electrical Engineer (FH). Heinrich joined ifb in 2015 leaving a career in science. Since then, he has worked as a consultant for different clients and participated in several ifb working groups on the analysis of new technologies, especially blockchain implementations and self-sovereign digital identities.

Matthias Kurfels Director at ifb SE, has been a consultant and trainer for banks, capital management companies and financial service providers since 2010. Before that, he was working for German savings banks and a savings bank association for more than 25 years.

His focus is on regulatory requirements for corporate and risk management, IT governance, internal audit and compliance. He is also head of ifb's internal working group for regulatory aspects of blockchain technology.

Dr. Sangmeng Li Senior Consultant at ifb SE, has primarily worked as a data scientist for quantitative risk management in the financial industry with a focus on data analysis, risk modeling and technical implementation. She received her doctorate in mathematics from the University of Münster, having conducted research on stochastic differential equation and Monte Carlo simulation as part of her Ph.D.

Volker Liermann Partner at ifb group, worked in the banking industry for over two decades, primarily focusing on financial risk management. Throughout his career, he has focused on developing integrated and comprehensive frameworks to help organizations correctly project risk at a strategic and tactical line of business and departmental level. He has also focused on developing frameworks to integrate stress testing and regulatory stress tests. In recent years, his focus has shifted to digitalization, machine learning and digital processes including improvements to classical financial and non-financial risk management. He has a background in economics and a degree in mathematics from the University of Bonn.



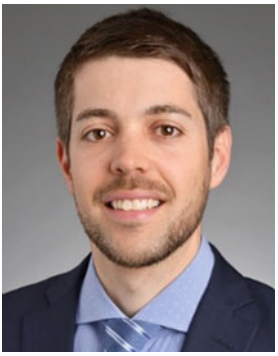
Michael Morawski Director, has worked at ifb group since 2008. He has conducted various projects in Germany and abroad mainly in the context of regulatory reporting and controlling, often involving ETL pipelines in systems like SAP SEM Banking, SAP Bank Analyzer, SAP BW, SAP HANA and lately also Hadoop. He focuses on the challenges posed by Big Data in the financial industry. Since 2018, he has coordinated the ifb internal Hadoop working group. Michael has a degree in biology from the University of Wuerzburg.

Georg Schmidt Managing Consultant, has worked at ifb group since October 2013. He started his career in the accounting sector as a developer for source data integration with SAP BW in Colombia and later in Peru. He also worked with a German automobile bank in the area of data transformation, working with the Hadoop tool stack, such as HDFS, Hive, Spark, Sqoop, Oozie, HBase and Kafka. Similarly, he developed a Spark data pipeline using Atlas as data catalog for an Austrian insurance company. Georg holds a master's degree in administration and informatics from the University of Potsdam.

Dr. Anne Seidlitz has been working as a consultant in the financial sector for more than four years. As part of the Business Intelligence & Data Governance Teams at ifb group, her focus is on the topics data integration, data warehousing, data modeling and data governance. She studied physics and holds a doctoral degree in this field from the Martin Luther University Halle-Wittenberg.

Claus Stegmann has as Co-CEO of ifb group—an international consulting firm—acquired extensive know-how over the last three decades in the financial industry regarding finance transformation, risk management and regulatory compliance. He is intensively engaged with the current challenges of the financial industry, which result from strong changes to customer behavior, a changing competitive environment and new technologies due to digitalization. He has also co-authored books on Stress Tests in Banks, Basel III as well as Digitalization in the Finance Industry, and graduated from Business School at the University of Passau, Germany.

Dr. Sascha Steltgens is a Consultant in ifb group's Information Management competence center. With a specialty in data integration, his work focuses on data delivery for all kinds of purposes in the financial sector. Since January 2019, his work has assisted banks and insurance companies with the implementation of state-of-the-art database technology. He studied molecular biology as well as biochemistry and holds a Ph.D. in biochemistry from Heinrich-Heine University Düsseldorf.



Ralph Steurer Director at ifb, has been working in the financial services industry for over ten years as architect and technical project manager. Over the years, he transitioned from SAP implementation projects (SAP Bank Analyzer and SAP HANA XSA) to software development, focusing on Apache Kafka, Java, NodeJS and ReactJS. His academic background includes a bachelor's degree in computer science from Zurich University of Applied Sciences.

Stefan Strube Senior Consultant, has been working at ifb group since 2018. He has supported well-known clients from the banking and insurance industry, most commonly dealing with both business and technology related topics and primarily focusing on business analysis, regulatory affairs and data analytics.

In addition, he acts as the product owner on the development of Modern Data Management, where he addresses the influence of the Digital Transformation as well as Open Banking on the development of German and European banks, and combines this with his work on modern business warehousing applications like SAP BW/4HANA.

He graduated with a Bachelor of Science from the University of Bonn and a Master of Science from the University of Göttingen.

Dr. Jeyakrishna Velauthapillai has been working in consulting since 2019 with a focus on SAP products such as SAP FPSL in combination with IFRS 17 topics. He is especially interested in the digital transformation of the insurance industry and its challenges. Additionally, he studied economics and mathematics, holds a doctoral degree in economics, and has experience in agent-based modeling and financial mathematics.

Bastian Wormuth has been working in consulting for over 17 years with a focus on information management and data analytics, most of the time in the financial industry. He leads the Business Intelligence & Data Governance Teams at ifb group. As an architect and data governance expert, he helps financial institutions to optimize data management processes and to manage data risk. Bastian obtained a degree in mathematics at the University of Technology Darmstadt and has published several papers on knowledge discovery and data analytics.

List of Figures

Data Lineage

Digitization and MongoDB—The Art of Possible

Graph Databases

Data Tiering Options with SAP HANA and Usage in a Hadoop Scenario

Kafka: Real-Time Streaming for the Finance Industry

Architecture Patterns—Batch and Real-Time Capabilities

Kafka—A Practical Implementation of Intraday Liquidity Risk Management

Data Sustainability—A Thorough Consideration

Special Data for Insurance Companies

Digital Identity Management—For Humans Only?

Overview Machine Learning and Deep Learning Frameworks