Multidimensional Array Data Management

Other titles in Foundations and Trends® in Databases

Modern Datalog Engines
Bas Ketsman and Paraschos Koutris
ISBN: 978-1-63828-042-2

Natural Language Interfaces to Data
Abdul Quamar, Vasilis Efthymiou, Chuan Lei and Fatma Özcan
ISBN: 978-1-63828-028-6

Trends in Explanations: Understanding and Debugging Data-driven Systems
Boris Glavic, Alexandra Meliou and Sudeepa Roy
ISBN: 978-1-68083-880-0

Differential Privacy for Databases
Joseph P. Near and Xi He
ISBN: 978-1-68083-850-3

Database Systems on GPUs
Johns Paul, Shengliang Lu and Bingsheng He
ISBN: 978-1-68083-848-0

Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases
Gerhard Weikum, Xin Luna Dong, Simon Razniewski and Fabian Suchanek
ISBN: 978-1-68083-836-7

Multidimensional Array Data Management

Florin Rusu
University of California Merced
frusu@ucmerced.edu

now
the essence of knowledge
Boston — Delft

Full text available at: http://dx.doi.org/10.1561/1900000069

Foundations and Trends® in Databases

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

F. Rusu. *Multidimensional Array Data Management*. Foundations and Trends® in Databases, vol. 12, no. 2-3, pp. 69–220, 2023.

ISBN: 978-1-63828-149-8
© 2023 F. Rusu

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, mechanical, photocopying, recording or otherwise, without prior written permission of the publishers.

Photocopying. In the USA: This journal is registered at the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by now Publishers Inc for users registered with the Copyright Clearance Center (CCC). The 'services' for users can be found on the internet at: www.copyright.com

For those organizations that have been granted a photocopy license, a separate system of payment has been arranged. Authorization does not extend to other kinds of copying, such as that for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. In the rest of the world: Permission to photocopy must be obtained from the copyright owner. Please apply to now Publishers Inc., PO Box 1024, Hanover, MA 02339, USA; Tel. +1 781 871 0245; www.nowpublishers.com; sales@nowpublishers.com

now Publishers Inc. has an exclusive license to publish this material worldwide. Permission to use this content must be obtained from the copyright license holder. Please apply to now Publishers, PO Box 179, 2600 AD Delft, The Netherlands, www.nowpublishers.com; e-mail: sales@nowpublishers.com

Full text available at: http://dx.doi.org/10.1561/1900000069

Foundations and Trends® in Databases

Volume 12, Issue 2-3, 2023

Editorial Board

Editorial Scope

Topics

Foundations and Trends® in Databases publishes survey and tutorial articles in the following topics:

- Data Models and Query Languages
- Query Processing and Optimization
- Storage, Access Methods, and Indexing
- Transaction Management, Concurrency Control and Recovery
- Deductive Databases
- Parallel and Distributed Database Systems
- Database Design and Tuning
- Metadata Management
- Object Management
- Trigger Processing and Active Databases
- Data Mining and OLAP
- Approximate and Interactive Query Processing
- Data Warehousing
- Adaptive Query Processing
- Data Stream Management
- Search and Query Integration
- XML and Semi-Structured Data
- Web Services and Middleware
- Data Integration and Exchange
- Private and Secure Data Management
- Peer-to-Peer, Sensornet and Mobile Data Management
- Scientific and Spatial Data Management
- Data Brokering and Publish/Subscribe
- Data Cleaning and Information Extraction
- Probabilistic Data Management

Information for Librarians

Foundations and Trends® in Databases, 2023, Volume 12, 4 issues. ISSN paper version 1931-7883. ISSN online version 1931-7891. Also available as a combined paper and online subscription.

Contents

Multidimensional Array Data Management

Florin Rusu

University of California Merced, USA; frusu@ucmerced.edu

ABSTRACT

Multidimensional arrays are a fundamental abstraction to represent data across scientific domains ranging from astronomy to genetics, medicine, business intelligence, and engineering. Arrays come under multiple shapes — from dense rasters to sparse data cubes and tensors — and have been studied extensively across many computing domains. In this survey, we provide a comprehensive guide for past, present, and future research in array data management from a database perspective. Unlike previous surveys that are limited to raster processing in the context of scientific data, we consider all types of arrays — rasters, data cubes, and tensors. We identify and analyze the most important research ideas on arrays proposed over time. We cover all data management aspects, from array algebras and query languages to storage strategies, execution techniques, and operator implementations. Moreover, we discuss which research ideas are adopted in real systems and how are they integrated in complete data processing pipelines. Finally, we compare arrays with the relational data model. The result is a thorough survey on array data management that should be consulted by anyone interested in this research topic — independent of experience level.

Florin Rusu (2023), “Multidimensional Array Data Management”, Foundations and Trends® in Databases: Vol. 12, No. 2-3, pp 69–220. DOI: 10.1561/1900000069.

1

Introduction

Multidimensional arrays are one of the fundamental computing abstractions to represent data across virtually all areas of science and engineering (Harris *et al.*, 2020) — and beyond. In science, spatio-temporal data acquired by sensors measuring environmental conditions or generated by simulations of physical phenomena are represented as 3- or 4-D dense arrays — also called *rasters* or *grids*. Concrete examples include spatial 3-D (x/y/z) arrays of Earth subsurface voxels, 3-D (x/y/t) time series of X-ray and fMRI medical images, and 4-D (x/y/z/t) optical or radio telescope signals in astronomy (Baumann *et al.*, 2021). In business analytics, *data cubes* aggregate statistical measures such as the mean, variance, and median across all the combinations of the values on a set of multiple — possibly hierarchical — dimensions (Gray *et al.*, 1996). For example, a retailer may want to compute the monthly average volume of sales for every city and every product category. Unlike the science rasters, the coordinates of a data cube do not necessarily have a strict ordering — they are categorical, not ordinal. Moreover, many entries in the data cube can be empty, resulting in a sparse array. In machine learning and artificial intelligence, highly-dimensional models are defined over features extracted from text and image data. For example, the text synthesis models applied in natural language processing consist of embeddings with billions of features (Brown *et al.*, 2020). These models are represented as 1-D vectors, 2-D matrices, and their multidimensional *tensor* generalizations. Machine learning training and prediction consist of a sequence of linear algebra operations between the model and the training/testing data — also represented as tensors.

Due to their ubiquity, multidimensional arrays — rasters, data cubes, and tensors — have been studied extensively across many areas of computer science — including compilers, programming languages, scientific and high-performance computing, graphics, machine learning, and databases. With the exception of databases, the vast majority of these studies are focused on the computational aspects of array processing — not the data management issues. Within the database field, the first "call" to extend the unordered set-based relational model with ordered rasters dates back to 1993 (Maier and Vance, 1993). Although rasters had initially spurred research interest, they had been overshadowed by data cubes upon their introduction in 1996 (Gray *et al.*, 1996). However, the era of "Big Data" from the late 2000s and early 2010s has renewed the interest in raster and sparse ordered arrays. The main driver has been the large volume of spatio-temporal data generated by scientific applications. This has led to the creation of the SciDB array database (Cudre-Mauroux *et al.*, 2009) as a collaboration between data management researchers and astronomers fostered by the XLDB suite of conferences. The success of deep neural network models in classifying objects from images and text has brought the spotlight on tensor processing in the late 2010s — which continues by the time of this writing.

In this work, we survey the research on multidimensional array data management — including rasters, data cubes, and tensors — from a database perspective. Thus, our focus is on work published in database conferences and journals. Nonetheless, we also include references to relevant work from other computing domains whenever necessary. Our definitive goal is to identify and analyze the most important research ideas on arrays proposed over time. We cover the full spectrum of data management, from array algebras and query languages to storage strategies, execution techniques, and operator implementations. Moreover, we discuss which research ideas are adopted in real systems and how are they integrated in complete data processing pipelines. Given that the unordered set-based relational data model is dominant across databases, we compare the differences to arrays at every step in the presentation.

The resulting survey aims to serve two main objectives. First, it summarizes concisely the most relevant work on multidimensional array data management by identifying the major research problems. Second, the survey organizes this material to provide an accurate perspective on the state-of-the-art and future directions in array processing. To the best of our knowledge, this is the first complete survey on array data management that includes rasters, data cubes, and tensors. Previous surveys are limited to raster processing in the context of scientific data. For example, the first survey on array storage and processing (Rusu and Cheng, 2013) does not consider data cubes and tensors. The VLDB 2021 tutorial on array DBMS (Zalipynis, 2021) discusses the design of the ChronosDB system for external raster processing.

The most recent survey on array databases (Baumann *et al.*, 2021) gives a thorough analytical and experimental comparison among several raster systems — nineteen systems are compared and four of them are experimentally benchmarked. While there is some unavoidable overlap between our work and this survey, their approach and take-away message are quite different. The focus of Baumann *et al.* (2021) is RasDaMan — the array database developed by the same authors. The presentation of the main concepts — including algebra and query language, storage, and processing — is centered on the solutions implemented in RasDaMan. Alternative solutions are only briefly referenced. The end message is that RasDaMan is the most feature-complete array database system available — which is true given its 30+ years of development. This work delves considerably deeper in each of the topics and covers more breadth. Our perspective on multidimensional array data management considers all types of arrays — rasters, data cubes, and tensors. The focus is first on methods and then on their realization in a particular system. Our end goal is to summarize all the methods in a systematic presentation and provide a thorough analysis. In summary, we view Baumann *et al.* (2021) as system-centric while this work is technique-centered.

This work surveys a large body of work on multidimensional arrays published over three decades and is organized as follows. We start with a theoretical formalization of arrays and their categorization in Section 2. The defining operations for every type of array are presented in Section 3. The next three sections follow the architecture of a data processing system — going top-down from the user interface to the execution internals. Array algebras and query languages are introduced in Section 4. Array storage techniques are presented in detail in Section 5. Execution strategies and array operators are discussed in Section 6. The implementation of these ideas and their integration in real systems are analyzed in Section 7. We conclude with a summary of the most relevant ideas and an outlook to future directions on array data management in Section 8.

References

Aberger, C., A. Lamb, K. Olukotun, and C. Re. (2018). “LevelHeaded: A Unified Engine for Business Intelligence and Linear Algebra Querying”. In: *Proceedings of 2018 IEEE ICDE International Conference on Data Engineering*. 449–460.

Agarwal, R. C., S. M. Balle, F. G. Gustavson, M. Joshi, and P. Palkar. (1995). “A Three-dimensional Approach to Parallel Matrix Multiplication”. *IBM Journal of Research and Development*. 39(5): 575–582.

Agrawal, R., A. Gupta, and S. Sarawagi. (1997). “Modeling Multidimensional Databases”. In: *Proceedings of 1997 IEEE ICDE International Conference on Data Engineering*. 232–243.

Alotaibi, R., D. Bursztyn, A. Deutsch, I. Manolescu, and S. Zampetakis. (2019). “Towards Scalable Hybrid Stores: Constraint-Based Rewriting to the Rescue”. In: *Proceedings of 2019 ACM SIGMOD International Conference on Management of Data*. 1660–1677.

Ballegooij, A. van, R. Cornacchia, A. P. de Vries, and M. Kersten. (2005). “Distribution Rules for Array Database Queries”. In: *Proceedings of 2005 DEXA International Conference on Database and Expert Systems Applications*. 55–64.

Ballegooij, A. R. van. (2004). “RAM: A Multidimensional Array DBMS”. In: *Proceedings of 2004 EDBT Extended Database Technology Workshops*. 154–165.

Barceló, P., N. Higuera, J. Pérez, and B. Subercaseaux. (2019). “Expressiveness of Matrix and Tensor Query Languages in Terms of ML Operators”. In: *Proceedings of 2019 DEEM International Workshop on Data Management for End-to-End Machine Learning.*

Baumann, P., A. Dehmel, P. Furtado, R. Ritsch, and N. Widmann. (1998). “The Multidimensional Database System RasDaMan”. In: *Proceedings of 1998 ACM SIGMOD International Conference on Management of Data*. 575–577.

Baumann, P. (1994). “On the Management of Multi-Dimensional Discrete Data”. *VLDB Journal (VLDBJ)*. 4(3): 401–444.

Baumann, P. (1999). “A Database Array Algebra for Spatio-Temporal Data and Beyond”. In: *Proceedings of 1999 NGITS International Workshop on Next Generation Information Technologies and Systems*. 76–93.

Baumann, P. and S. Holsten. (2011). “A Comparative Analysis of Array Models for Databases”. In: *Proceedings of 2011 FGIT-DTA/BSBT*. 80–89.

Baumann, P. and V. Merticariu. (2015). “On the Efficient Evaluation of Array Joins”. In: *Proceedings of 2015 IEEE BIG DATA International Conference on Big Data*. 2046–2055.

Baumann, P., D. Misev, V. Merticariu, and B. P. Huu. (2021). “Array Databases: Concepts, Standards, Implementations”. *Journal of Big Data*. 8(1).

Beryoza, D., M. Campbell, C. Cardorelle, T. Creasey, D. Cushing, V. D. Silva, S. David, A. Hagleitner, I. Henderson, D. Howell, I. Kozine, P. Prieto, P. Thompson, J. Vazquez, and Y. Zhang. (2015). *IBM Cognos Dynamic Cubes*. IBM Redbooks.

Blanas, S., J. M. Patel, V. Ercegovac, J. Rao, E. J. Shekita, and Y. Tian. (2010). “A Comparison of Join Algorithms for Log Processing in MapReduce”. In: *Proceedings of 2010 ACM SIGMOD International Conference on Management of Data*. 975–986.

Blanas, S., K. Wu, S. Byna, B. Dong, and A. Shoshani. (2014). “Parallel Data Analysis Directly on Scientific File Formats”. In: *Proceedings of 2014 ACM SIGMOD International Conference on Management of Data*. 385–396.

Boehm, M., M. W. Dusenberry, D. Eriksson, A. V. Evfimievski, F. M. Manshadi, N. Pansare, B. Reinwald, F. R. Reiss, P. Sen, A. C. Surve, and S. Tatikonda. (2016). "SystemML: Declarative Machine Learning on Spark". *PVLDB*. 9(13): 1425–1436.

Boehm, M., B. Reinwald, D. Hutchison, P. Sen, A. V. Evfimievski, and N. Pansare. (2018). "On Optimizing Operator Fusion Plans for Large-Scale Machine Learning in SystemML". *PVLDB*. 11(12): 1755–1768.

Boncz, P., M. Zukowski, and N. Nes. (2005). "MonetDB/X100: Hyper-Pipelining Query Execution". In: *Proceedings of 2005 CIDR Conference on Innovative Database Research*. 225–237.

Brijder, R., F. Geerts, J. V. D. Bussche, and T. Weerwag. (2019). "On the Expressive Power of Query Languages for Matrices". *ACM Transactions on Database Systems (TODS)*. 44(4).

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. (2020). "Language Models are Few-Shot Learners". *CoRR*. arXiv:2005.14165.

Buck, J., N. Watkins, G. Levin, A. Crume, K. Ioannidou, S. Brandt, C. Maltzahn, and N. Polyzotis. (2012). "SIDR: Efficient Structure-Aware Intelligent Data Routing in SciHadoop". *Tech. rep.* No. UCSC-TR-SOE-12-08. UC Santa Cruz.

Buck, J. B., N. Watkins, J. LeFevre, K. Ioannidou, C. Maltzahn, N. Polyzotis, and S. Brandt. (2011). "SciHadoop: Array-based Query Processing in Hadoop". In: *Proceedings of 2011 SC International Conference for High Performance Computing, Networking, Storage and Analysis*. 66:1–66:11.

Cabibbo, L. and R. Torlone. (1998). "A Logical Approach to Multidimensional Databases". In: *Proceedings of 1998 EDBT Extended Database Technology Workshops*. 183–197.

Cao, B. and A. Badia. (2007). "SQL Query Optimization Through Nested Relational Algebra". *ACM Transactions on Database Systems (TODS)*. 32(3).

Chang, C., A. Acharya, A. Sussman, and J. H. Saltz. (1998). “T2: A Customizable Parallel Database for Multi-Dimensional Data”. *SIGMOD Rec.* 27(1): 58–66.

Chang, C., B. Moon, A. Acharya, C. Shock, A. Sussman, and J. H. Saltz. (1997). “Titan: A High-Performance Remote Sensing Database”. In: *Proceedings of 1997 IEEE ICDE International Conference on Data Engineering*. 375–384.

Chaudhuri, S. and U. Dayal. (1997). “An Overview of Data Warehousing and OLAP Technology”. *ACM SIGMOD Record*. 26(1): 65–74.

Chaudhuri, S., U. Dayal, and V. Narasayya. (2011). “An Overview of Business Intelligence Technology”. *Commun. ACM*. 54(8): 88–98.

Cheng, Y., C. Qin, and F. Rusu. (2012). “GLADE: Big Data Analytics Made Easy”. In: *Proceedings of 2012 ACM SIGMOD International Conference on Management of Data*. 697–700.

Cheng, Y. and F. Rusu. (2014). “Formal Representation of the SS-DB Benchmark and Experimental Evaluation in EXTASCID”. *Distributed and Parallel Databases*.

Choi, D., C.-S. Park, and Y. D. Chung. (2019). “Progressive Top-k Subarray Query Processing in Array Databases”. *PVLDB*. 12(9): 989–1001.

Choi, J., J. J. Dongarra, R. Pozo, and D. W. Walker. (1992). “ScaLAPACK: A Scalable Linear Algebra Library for Distributed Memory Concurrent Computers”. In: *Proceedings of 1992 Symposium on the Frontiers of Massively Parallel Computation*. 120–127.

CMCC Foundation. (2022a). “Ophidia Big Data Code Repository”. URL: https://github.com/OphidiaBigData.

CMCC Foundation. (2022b). “Ophidia Project”. URL: https://ophidia.cmcc.it/.

Codd, E. (1970). “A Relational Model for Large Shared Data Banks”. *Comm. ACM*. 13(6): 377–387.

Cornacchia, R., S. Héman, M. Zukowski, A. P. de Vries, and P. Boncz. (2008). “Flexible and Efficient IR using Array Databases”. *VLDB Journal (VLDBJ)*. 17: 151–168.

Cudre-Mauroux, P., H. Kimura, K.-T. Lim, J. Rogers, R. Simakov, E. Soroush, P. Velikhov, D. L. Wang, M. Balazinska, J. Becla, D. DeWitt, B. Heath, D. Maier, S. Madden, J. Patel, M. Stonebraker, and S. Zdonik. (2009). "A Demonstration of SciDB: A Science-Oriented DBMS". *PVLDB*. 2(2): 1534–1537.

Cudre-Mauroux, P., H. Kimura, K.-T. Lim, J. Rogers, S. Madden, M. Stonebraker, S. B. Zdonik, and P. G. Brown. (2010). "SS-DB: A Standard Science DBMS Benchmark". URL: http://www.xldb.org/science-benchmark/.

D'silva, J. V., F. De Moor, and B. Kemme. (2018). "AIDA: Abstraction for Advanced in-Database Analytics". *PVLDB*. 11(11): 1400–1413.

Datta, K., M. Murphy, V. Volkov, S. Williams, J. Carter, L. Oliker, D. Patterson, J. Shalf, and K. Yelick. (2008). "Stencil Computation Optimization and Auto-tuning on State-of-the-art Multicore Architectures". In: *Proceedings of 2008 SC International Conference for High Performance Computing, Networking, Storage and Analysis*. 1–12.

Dean, J. and S. Ghemawat. (2008). "MapReduce: Simplified Data Processing on Large Clusters". *Commun. ACM*. 51(1): 107–113.

Dewitt, D. J., S. Ghandeharizadeh, D. A. Schneider, A. Bricker, H. Hsiao, and R. Rasmussen. (1990). "The Gamma Database Machine Project". *IEEE Transactions on Knowledge and Data Engineering (TKDE)*. 2(1): 44–62.

DeWitt, D. J. and J. Gray. (1991). "Parallel Database Systems: The Future of Database Processing or a Passing Fad?" *SIGMOD Rec.* 19.

Dolmatova, O., N. Augsten, and M. H. Böhlen. (2020). "A Relational Matrix Algebra and Its Implementation in a Column Store". In: *Proceedings of 2020 ACM SIGMOD International Conference on Management of Data*. 2573–2587.

Dong, B., K. Wu, S. Byna, J. Liu, W. Zhao, and F. Rusu. (2017). "ArrayUDF: User-Defined Scientific Data Analysis on Arrays". In: *Proceedings of 2017 ACM HPDC International Symposium on High-Performance Parallel and Distributed Computing.*

Dongarra, J. and R. Schreiber. (1990). "Automatic Blocking of Nested Loops". *Tech. rep.* University of Tennessee.

Duggan, J., A. J. Elmore, M. Stonebraker, M. Balazinska, B. Howe, J. Kepner, S. Madden, D. Maier, T. Mattson, and S. Zdonik. (2015a). "The BigDAWG Polystore System". *ACM SIGMOD Record*. 44(2): 11–16.

Duggan, J., O. Papaemmanouil, L. Battle, and M. Stonebraker. (2015b). "Skew-Aware Join Optimization for Array Databases". In: *Proceedings of 2015 ACM SIGMOD International Conference on Management of Data*. 123–135.

Duggan, J. and M. Stonebraker. (2014). "Incremental Elasticity for Array Databases". In: *Proceedings of 2014 ACM SIGMOD International Conference on Management of Data*. 409–420.

Eigen Development Team. (2010). "Eigen". URL: http://eigen.tuxfamily.org.

Elgohary, A., M. Boehm, P. J. Haas, F. R. Reiss, and B. Reinwald. (2016). "Compressed Linear Algebra for Large-Scale Machine Learning". *PVLDB*. 9(12): 960–971.

Faloutsos, C. and P. Bhagwat. (1993). "Declustering Using Fractals". In: *Proceedings of 1993 International Conference on Parallel and Distributed Information Systems*. 18–25.

Furtado, P. and P. Baumann. (1999). "Storage of Multidimensional Arrays Based on Arbitrary Tiling". In: *Proceedings of 1999 IEEE ICDE International Conference on Data Engineering*. 480–489.

Gao, Z. J., S. Luo, L. L. Perez, and C. Jermaine. (2017). "The BUDS Language for Distributed Bayesian Machine Learning". In: *Proceedings of 2017 ACM SIGMOD International Conference on Management of Data*. 961–976.

Ge, T., D. Grabiner, and S. Zdonik. (2011). "Monte Carlo Query Processing of Uncertain Multidimensional Array Data". In: *Proceedings of 2011 IEEE ICDE International Conference on Data Engineering*. 936–947.

Ge, T. and S. Zdonik. (2010). "A*-Tree: A Structure for Storage and Modeling of Uncertain Multidimensional Arrays". *PVLDB*. 3(1): 964–974.

Goil, S. and A. N. Choudhary. (1997). "Sparse Data Storage Schemes for Multidimensional Data for OLAP and Data Mining". *Tech. rep.* No. CPDC-TR-9801-005. Northwestern University.

Goto, K. and R. van de Geijn. (2008). “Anatomy of High-performance Matrix Multiplication”. *ACM Transactions on Mathematical Software (TOMS)*. 34(3).

Goto, K. and R. van de Geijn. (2009). “High-performance Implementation of the Level-3 BLAS”. *ACM Transactions on Mathematical Software (TOMS)*. 35(1).

Gray, J., A. Bosworth, A. Layman, and H. Pirahesh. (1996). “Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Total”. In: *Proceedings of 1996 IEEE ICDE International Conference on Data Engineering*. 152–159.

Gu, R., Y. Tang, C. Tian, H. Zhou, G. Li, X. Zheng, and Y. Huang. (2017). “Improving Execution Concurrency of Large-Scale Matrix Multiplication on Distributed Data-Parallel Platforms”. *IEEE Transactions on Parallel and Distributed Systems (TPDS)*. 28(9): 2539–2552.

Guo, H. (2021). *What Are Tensors Exactly?* World Scientific.

Guttman, A. (1984). “R-trees: A Dynamic Index Structure for Spatial Searching”. In: *Proceedings of 1984 ACM SIGMOD International Conference on Management of Data*. 47–57.

Gyssens, M. and L. V. Lakshmanan. (1997). “A Foundation for Multi-Dimensional Databases”. In: *Proceedings of 1997 VLDB International Conference on Very Large Data Bases*. 106–115.

Hadoop Development Team. (2020). “Hadoop”. URL: http://hadoop.apache.org/.

Han, D., Y.-M. Nam, J. Lee, K. Park, H. Kim, and M.-S. Kim. (2019). “DistME: A Fast and Elastic Distributed Matrix Computation Engine using GPUs”. In: *Proceedings of 2019 ACM SIGMOD International Conference on Management of Data*. 759–774.

Harinarayan, V., A. Rajaraman, and J. D. Ullman. (1996). “Implementing Data Cubes Efficiently”. In: *Proceedings of 1996 ACM SIGMOD International Conference on Management of Data*. 205–216.

Harris, C. R., K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gerard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. (2020). "Array Programming with NumPy". *Nature*. 585(7825): 357–362.

Hellerstein, J. M., C. Ré, F. Schoppmann, D. Z. Wang, E. Fratkin, A. Gorajek, K. S. Ng, C. Welton, X. Feng, K. Li, and A. Kumar. (2012). "The MADlib Analytics Library: Or MAD Skills, the SQL". *PVLDB*. 5(12): 1700–1711.

Hong, C., A. Sukumaran-Rajam, I. Nisa, K. Singh, and P. Sadayappan. (2019). "Adaptive Sparse Tiling for Sparse Matrix Multiplication". In: *Proceedings of 2019 PPoPP Symposium on Principles and Practice of Parallel Programming*. 300–314.

Horlova, O., A. Kaitoua, and S. Ceri. (2020). "Array-based Data Management for Genomics". In: *Proceedings of 2020 IEEE ICDE International Conference on Data Engineering*. 109–120.

Howe, B. and D. Maier. (2004). "Algebraic Manipulation of Scientific Datasets". In: *Proceedings of 2004 VLDB International Conference on Very Large Data Bases*. 924–935.

Hoyer, S. and J. Hamman. (2017). "xarray: N-D labeled Arrays and Datasets in Python". *Journal of Open Research Software*. 5(1).

Huang, B., S. Babu, and J. Yang. (2013). "Cumulon: Optimizing Statistical Data Analysis in the Cloud". In: *Proceedings of 2013 ACM SIGMOD International Conference on Management of Data*. 1–12.

Hutchison, D., B. Howe, and D. Suciu. (2017). "LaraDB: A Minimalist Kernel for Linear and Relational Algebra Computation". In: *Proceedings of 2017 ACM SIGMOD BeyondMR Workshop on Algorithms and Systems for MapReduce and Beyond*.

Idreos, S., F. Groffen, N. Nes, S. Manegold, K. S. Mullender, and M. L. Kersten. (2012). "MonetDB: Two Decades of Research in Column-Oriented Database Architectures". *IEEE Data Eng. Bull.* 35(1): 40–45.

Irigoin, F. and R. Triolet. (1988). "Supernode Partitioning". In: *Proceedings of 1988 ACM POPL Symposium on Principles of Programming Languages*. 319–329.

Ivanova, M., M. L. Kersten, and S. Manegold. (2012). "Data Vaults: A Symbiosis between Database Technology and Scientific File Repositories". In: *Proceedings of 2012 SSDBM International Conference on Scientific and Statistical Database Management*. 485–494.

Jaeschke, G. and H. J. Schek. (1982). "Remarks on the Algebra of Non First Normal Form Relations". In: *Proceedings of 1982 PODS Symposium on Principles of Database Systems*. 124–138.

Jagadish, H. (1990). "Linear Clustering of Objects with Multiple Attributes". In: *Proceedings of 1990 ACM SIGMOD International Conference on Management of Data*. 332–342.

Jankov, D., B. Yuan, S. Luo, and C. Jermaine. (2021). "Distributed Numerical and Machine Learning Computations via Two-Phase Execution of Aggregated Join Trees". *PVLDB*. 14(7): 1228–1240.

Jordan, H., P. Thoman, J. Durillo, S. Pellegrini, P. Gschwandtner, T. Fahringer, and H. Moritsch. (2012). "A Multi-objective Auto-tuning Framework for Parallel Codes". In: *Proceedings of 2012 SC International Conference for High Performance Computing, Networking, Storage and Analysis*. 10:1–10:12.

Kernert, D., W. Lehner, and F. Kohler. (2016). "Topology-aware Optimization of Big Sparse Matrices and Matrix Multiplications on Main-memory Systems". In: *Proceedings of 2016 IEEE ICDE International Conference on Data Engineering*. 823–834.

Kersten, M. L., Y. Zhang, M. Ivanova, and N. Nes. (2011). "SciQL, A Query Language for Science Applications". In: *Proceedings of 2011 AD EDBT/ICDT Array Databases Workshop*. 1–12.

Khamis, M. A., H. Q. Ngo, X. Nguyen, D. Olteanu, and M. Schleich. (2018). "AC/DC: In-Database Learning Thunderstruck". In: *Proceedings of 2018 DEEM International Workshop on Data Management for End-to-End Machine Learning*.

Kim, J., A. Sukumaran-Rajam, V. Thumma, S. Krishnamoorthy, A. Panyala, L. Pouchet, A. Rountev, and P. Sadayappan. (2019). "A Code Generator for High-Performance Tensor Contractions on GPUs". In: *Proceedings of 2019 IEEE/ACM CGO International Symposium on Code Generation and Optimization*. 85–95.

Kim, M. (2014). *TensorDB and Tensor-Relational Model (TRM) for Efficient Tensor-Relational Operations, Ph.D. Thesis*. Arizona State University.

Kim, M. and K. S. Candan. (2011). "Approximate Tensor Decomposition within a Tensor-Relational Algebraic Framework". In: *Proceedings of 2011 ACM CIKM International Conference on Information and Knowledge Management*. 1737–1742.

Kim, M. and K. S. Candan. (2014). "Efficient Static and Dynamic In-Database Tensor Decompositions on Chunk-Based Array Stores". In: *Proceedings of 2014 ACM CIKM International Conference on Information and Knowledge Management*. 969–978.

Kim, S., B. Kim, and B. Moon. (2021). "Spangle: A Distributed In-Memory Processing System for Large-Scale Arrays". In: *Proceedings of 2021 IEEE ICDE International Conference on Data Engineering*. 1799–1810.

Kisuki, T., P. M. W. Knijnenburg, and M. F. P. O'Boyle. (2000). "Combined Selection of Tile Sizes and Unroll Factors Using Iterative Compilation". In: *Proceedings of 2000 PACT International Conference on Parallel Architectures and Compilation Techniques*. 237–248.

Kolda, T. G. and B. W. Bader. (2009). "Tensor Decompositions and Applications". *SIAM Rev.* 51(3): 455–500.

Koutsoukos, D., S. Nakandala, K. Karanasos, K. Saur, G. Alonso, and M. Interlandi. (2021). "Tensors: An Abstraction for General Data Processing". *PVLDB*. 14(10): 1797–1804.

Kunft, A., A. Alexandrov, A. Katsifodimos, and V. Markl. (2016). "Bridging the Gap: Towards Optimization across Linear and Relational Algebra". In: *Proceedings of 2016 ACM SIGMOD BeyondMR Workshop on Algorithms and Systems for MapReduce and Beyond*.

Kylin Development Team. (2022). "Kylin". URL: https://kylin.apache.org/.

Laboratory for Foundations of Computer Science at the University of Edinburgh. (2008). “The Standard ML Language”. URL: http://www.lfcs.inf.ed.ac.uk/software/ML/.

Laue, S., M. Mitterreiter, and J. Giesen. (2020). “A Simple and Efficient Tensor Calculus”. In: *Proceedings of 2020 AAAI Conference on Artificial Intelligence*. 4527–4534.

Lerner, A. and D. Shasha. (2003). “AQuery: Query Language for Ordered Data, Optimization Techniques, and Experiments”. In: *Proceedings of 2003 VLDB International Conference on Very Large Data Bases*. 345–356.

Leung, A., N. Vasilache, B. Meister, M. Baskaran, D. Wohlford, C. Bastoul, and R. Lethin. (2010). “A Mapping Path for Multi-GPGPU Accelerated Computers from a Portable High Level Programming Abstraction”. In: *Proceedings of 2010 Workshop on General-Purpose Computation on Graphics Processing Units*. 51–61.

Li, F., L. Chen, Y. Zeng, A. Kumar, X. Wu, J. F. Naughton, and J. M. Patel. (2019a). “Tuple-Oriented Compression for Large-Scale Mini-Batch Stochastic Gradient Descent”. In: *Proceedings of 2019 ACM SIGMOD International Conference on Management of Data*. 1517–1534.

Li, R., W. Gatterbauer, and M. Riedewald. (2020). “Near-Optimal Distributed Band-Joins through Recursive Partitioning”. In: *Proceedings of 2020 ACM SIGMOD International Conference on Management of Data*. 2375–2390.

Li, X., Y. Liang, S. Yan, L. Jia, and Y. Li. (2019b). “A Coordinated Tiling and Batching Framework for Efficient GEMM on GPUs”. In: *Proceedings of 2019 PPoPP Symposium on Principles and Practice of Parallel Programming*. 229–241.

Libkin, L., R. Machlin, and L. Wong. (1996). “A Query Language for Multidimensional Arrays: Design, Implementation, and Optimization Techniques”. In: *Proceedings of 1996 ACM SIGMOD International Conference on Management of Data*. 228–239.

Lim, K.-T., D. Maier, J. Becla, M. Kersten, Y. Zhang, and M. Stonebraker. (2012). “Array QL Syntax”. URL: http://www.xldb.org/wp-content/uploads/2012/09/ArrayQL-Draft-4.pdf.

Lippmeier, B. and G. Keller. (2011). “Efficient Parallel Stencil Convolution in Haskell”. In: *Proceedings of 2011 Haskell ACM Symposium on Haskell*. 59–70.

Liu, D.-R. and S. Shekhar. (1995). “A Similarity Graph-based Approach to Declustering Problems and Its Application towards Parallelizing Grid Files”. In: *Proceedings of 1995 IEEE ICDE International Conference on Data Engineering*. 373–381.

Lowenthal, D. K. (2000). “Accurately Selecting Block Size at Runtime in Pipelined Parallel Programs”. *Int. J. Parallel Program.* 28(3): 245–274.

Luo, S. (2020). *Automatic Matrix Format Exploration for Large Scale Linear Algebra, Ph.D. Thesis*. Rice University.

Luo, S., Z. J. Gao, M. N. Gubanov, L. L. Perez, and C. M. Jermaine. (2017). “Scalable Linear Algebra on a Relational Database System”. In: *Proceedings of 2017 IEEE ICDE International Conference on Data Engineering*. 523–534.

Luo, S., D. Jankov, B. Yuan, and C. Jermaine. (2021). “Automatic Optimization of Matrix Implementations for Distributed Machine Learning and Linear Algebra”. In: *Proceedings of 2021 ACM SIGMOD International Conference on Management of Data*. 1222–1234.

Lustosa, H. and F. Porto. (2019). “SAVIME: A Multidimensional System for the Analysis and Visualization of Simulation Data”. *CoRR*. arXiv:1903.02949v2.

MADlib Development Team. (2022). “MADlib”. URL: https://madlib.apache.org/.

Maier, D. (2012). “ArrayQL Algebra: version 3”. URL: http://www.xldb.org/wp-content/uploads/2012/09/ArrayQL_Algebra_v3+.pdf.

Maier, D. and B. Vance. (1993). “A Call to Order”. In: *Proceedings of 1993 PODS Symposium on Principles of Database Systems*. 1–16.

Marathe, A. P. and K. Salem. (2002). “Query Processing Techniques for Arrays”. *VLDB Journal (VLDBJ)*. 11(1): 68–91.

Maruyama, N., T. Nomura, K. Sato, and S. Matsuoka. (2011). “Physis: An Implicitly Parallel Programming Model for Stencil Computations on Large-Scale GPU-Accelerated Supercomputers”. In: *Proceedings of 2011 SC International Conference for High Performance Computing, Networking, Storage and Analysis*. 11:1–11:12.

Matthews, D. A. (2016). “High-Performance Tensor Contraction without Transposition”. *CoRR*. arXiv:1607.00291v4.

Milner, R., M. Tofte, R. Harper, and D. MacQueen. (1997). *The Definition of Standard ML (revised)*. MIT Press.

Milo, T. and E. Altshuler. (2016). “An Efficient MapReduce Cube Algorithm for Varied Data Distributions”. In: *Proceedings of 2016 ACM SIGMOD International Conference on Management of Data*. 1151–1165.

Misev, D. and P. Baumann. (2014). “Extending the SQL Array Concept to Support Scientific Analytics”. In: *Proceedings of 2014 SSDBM International Conference on Scientific and Statistical Database Management*.

Moon, B., A. Acharya, and J. Saltz. (1996). “Study of Scalable Declustering Algorithms for Parallel Grid Files”. In: *Proceedings of 1996 Parallel Processing Symposium*. 434–440.

Moon, B. and J. H. Saltz. (1998). “Scalability Analysis of Declustering Methods for Multidimensional Range Queries”. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*. 10(2): 310–327.

Morfonios, K., S. Konakas, Y. Ioannidis, and N. Kotsis. (2007). “ROLAP Implementations of the Data Cube”. *ACM Comput. Surv.* 39(4).

MPI Forum. (2022). “Message Passing Interface (MPI)”. URL: https://www.mpi-forum.org/.

Nandi, A., C. Yu, P. Bohannon, and R. Ramakrishnan. (2012). “Data Cube Materialization and Mining over MapReduce”. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*. 24(10): 1747–1759.

Nikolopoulos, D. (2004). “Dynamic Tiling for Effective Use of Shared Caches on Multithreaded Processors”. *International Journal of High Performance Computing and Networking*. 2(1): 22–35.

NumPy Development Team. (2022). “NumPy”. URL: https://numpy.org/.

NVIDIA. (2022). “cuBLAS”. URL: https://docs.nvidia.com/cuda/cublas/.

O’Gorman, L., M. J. Sammon, and M. Seul. (2008). *Practical Algorithms for Image Analysis, 2nd edition*. Cambridge University Press.

Open Geospatial Consortium. (2022). "GeoTIFF Standard". URL: https://www.ogc.org/standards/geotiff.

Oracle. (2022a). "Essbase". URL: https://docs.oracle.com/en/database/other-databases/essbase/index.html.

Oracle. (2022b). "Spatial GeoRaster". URL: https://docs.oracle.com/en/database/oracle/oracle-database/21/geors/index.html.

Otoo, E. J., D. Rotem, and S. Seshadri. (2007). "Optimal Chunking of Large Multidimensional Arrays for Data Warehousing". In: *Proceedings of 2007 ACM DOLAP International Workshop on Data Warehousing and OLAP*. 25–32.

Ozsoyoglu, G., Z. M. Ozsoyoglu, and V. Matos. (1987). "Extending Relational Algebra and Relational Calculus with Set-valued Attributes and Aggregate Functions". *ACM Transactions on Database Systems (TODS)*. 12(4): 566–592.

Papadopoulos, S., K. Datta, S. Madden, and T. G. Mattson. (2016). "The TileDB Array Data Storage Manager". *PVLDB*. 10(4): 349–360.

Paradigm4. (2022a). "SciDB". URL: https://github.com/Paradigm4/SciDB.

Paradigm4. (2022b). "SciDB Documentation". URL: https://paradigm4.atlassian.net/wiki/spaces/scidb/overview?homepageId=2694289094.

Pedersen, T. B., C. S. Jensen, and C. E. Dyreson. (2001). "A Foundation for Capturing and Querying Complex Multidimensional Data". *Information Systems*. 26(5): 383–423.

Peng, L. and Y. Diao. (2015). "Supporting Data Uncertainty in Array Databases". In: *Proceedings of 2015 ACM SIGMOD International Conference on Management of Data*. 545–560.

Pentaho. (2022). "Mondrian OLAP Server". URL: https://mondrian.pentaho.com/documentation/index.php.

Polychroniou, O., R. Sen, and K. A. Ross. (2014). "Track Join: Distributed Joins with Minimal Network Traffic". In: *Proceedings of 2014 ACM SIGMOD International Conference on Management of Data*. 1483–1494.

Prabhakar, S., K. Abdel-Ghaffar, D. Agrawal, and A. E. Abbadi. (1998). "Cyclic Allocation of Two-Dimensional Data". In: *Proceedings of 1998 IEEE ICDE International Conference on Data Engineering.* 94–101.

PyTorch Development Team. (2022). "PyTorch". URL: https://pytorch.org/.

Qin, C. and F. Rusu. (2015). "Speculative Approximations for Terascale Distributed Gradient Descent Optimization". In: *Proceedings of 2015 ACM SIGMOD DanaC Workshop on Data Analytics in the Cloud.*

RasDaMan Development Team. (2022). "RasDaMan". URL: http://rasdaman.org/.

Renganarayana, L. and S. Rajopadhye. (2008). "Positivity, Posynomials and Tile Size Selection". In: *Proceedings of 2008 ACM/IEEE SC Conference on Supercomputing.*

Ritter, G., J. Wilson, and J. Davidson. (1990). "Image Algebra: An Overview". *Computer Vision, Graphics, and Image Processing.* 49(1): 297–331.

Rusu, F. and Y. Cheng. (2013). "A Survey on Array Storage, Query Languages, and Systems". *CoRR.* arXiv:1302.0103.

Salton, G., A. Wong, and C. S. Yang. (1975). "A Vector Space Model for Automatic Indexing". *Commun. ACM.* 18(11): 613–620.

Sarawagi, S. and M. Stonebraker. (1994). "Efficient Organization of Large Multidimensional Arrays". In: *Proceedings of 1994 IEEE ICDE International Conference on Data Engineering.* 328–336.

ScaLAPACK Development Team. (2022). "Scalable Linear Algebra PACKage (ScaLAPACK)". URL: http://www.netlib.org/scalapack/.

Seering, A., P. Cudre-Mauroux, S. Madden, and M. Stonebraker. (2012). "Efficient Versioning for Scientific Array Databases". In: *Proceedings of 2012 IEEE ICDE International Conference on Data Engineering.* 1013–1024.

Seo, S., E. J. Yoon, J. Kim, S. Jin, J.-S. Kim, and S. Maeng. (2010). "HAMA: An Efficient Matrix Computation with the MapReduce Framework". In: *Proceedings of 2010 IEEE CLOUDCOM International Conference on Cloud Computing Technology and Science.* 721–726.

Shao, Z., J. Han, and D. Xin. (2004). "MM-Cubing: Computing Iceberg Cubes by Factorizing the Lattice Space". In: *Proceedings of 2004 SSDBM International Conference on Scientific and Statistical Database Management*.

Shi, Y., U. N. Niranjan, A. Anandkumar, and C. Cecka. (2016). "Tensor Contractions with Extended BLAS Kernels on CPU and GPU". *CoRR*. arXiv:1606.05696.

Shoshani, A. (1997). "OLAP and Statistical Databases: Similarities and Differences". In: *Proceedings of 1997 PODS Symposium on Principles of Database Systems*. 185–196.

Sommer, J., M. Boehm, A. V. Evfimievski, B. Reinwald, and P. J. Haas. (2019). "MNC: Structure-Exploiting Sparsity Estimation for Matrix Expressions". In: *Proceedings of 2019 ACM SIGMOD International Conference on Management of Data*. 1607–1623.

Song, J., H. V. Jagadish, and G. Alter. (2021). "SDTA: An Algebra for Statistical Data Transformation". In: *Proceedings of 2021 SSDBM International Conference on Scientific and Statistical Database Management*. 109–120.

Soroush, E. and M. Balazinska. (2013). "Time Travel in a Scientific Array Database". In: *Proceedings of 2013 IEEE ICDE International Conference on Data Engineering*.

Soroush, E., M. Balazinska, and D. L. Wang. (2011). "ArrayStore: A Storage Manager for Complex Parallel Array Processing". In: *Proceedings of 2011 ACM SIGMOD International Conference on Management of Data*. 253–264.

Springer, P. and P. Bientinesi. (2018). "Design of a High-Performance GEMM-like Tensor–Tensor Multiplication". *ACM Transactions on Mathematical Software (TOMS)*. 44(3).

Springer, P. and C.-H. Yu. (2019). "cuTENSOR: High-Performance CUDA Tensor Primitives". URL: https://developer.nvidia.com/cutensor.

Stonebraker, M., P. Brown, A. Poliakov, and S. Raman. (2011). "The Architecture of SciDB". In: *Proceedings of 2011 SSDBM International Conference on Scientific and Statistical Database Management*. 1–16.

Szalay, A. S. (2008). “The Sloan Digital Sky Survey and Beyond”. *SIGMOD Rec.* 37(2): 61–66.

TensorFlow Development Team. (2022). “TensorFlow”. URL: https://www.tensorflow.org/.

The Dask Development Team. (2022). “Dask Arrays”. URL: https://docs.dask.org/en/latest/array.html.

The FITS Support Office. (2022). “Flexible Image Transport System (FITS)”. URL: https://fits.gsfc.nasa.gov/fits_home.html.

The GDAL Development Team. (2022). “Geospatial Data Abstraction Library (GDAL)”. URL: https://gdal.org/.

The HDF5 Group. (2020). “HDF5”. URL: http://www.hdfgroup.org/HDF5/.

The netCDF Development Team. (2022). “netCDF Operators (NCO)”. URL: http://nco.sourceforge.net/.

The PostGIS Development Team. (2022). “PostGIS Raster Data Management and Applications”. URL: http://postgis.net/docs/manual-dev/using_raster_dataman.html.

The PostgreSQL Development Team. (2020). “PostgreSQL”. URL: http://www.postgresql.org/.

Thomas, A. and A. Kumar. (2018). “A Comparative Evaluation of Systems for Scalable Linear Algebra-based Analytics”. *PVLDB*. 11(13): 2168–2182.

TileDB, Inc. (2022a). “TileDB”. URL: https://github.com/TileDB-Inc.

TileDB, Inc. (2022b). “TileDB Documentation”. URL: https://docs.tiledb.com/.

Tomlin, C. D. (1990). *Geographic Information Systems and Cartographic Modelling*. Prentice Hall.

Torlone, R. (2003). “Multidimensional Databases”. In: IGI Global. Chap. Conceptual Multidimensional Models.

UniData. (2022). “Network Common Data Form (NetCDF)”. URL: https://www.unidata.ucar.edu/software/netcdf/.

Vasilache, N., O. Zinenko, T. Theodoridis, P. Goyal, Z. DeVito, W. S. Moses, S. Verdoolaege, A. Adams, and A. Cohen. (2018). “Tensor Comprehensions: Framework-Agnostic High-Performance Machine Learning Abstractions”. *CoRR*. arXiv:1802.04730.

Vassiliadis, P. (1998). "Modeling Multidimensional Databases, Cubes and Cube Operations". In: *Proceedings of 1998 SSDBM International Conference on Scientific and Statistical Database Management*. 53–62.

Vassiliadis, P. and T. Sellis. (1999). "A Survey of Logical Models for OLAP Databases". *ACM SIGMOD Record*. 28(4): 64–69.

Wang, Y., A. Nandi, and G. Agrawal. (2014). "SAGA: Array Storage as a DB with Support for Structural Aggregations". In: *Proceedings of 2014 SSDBM International Conference on Scientific and Statistical Database Management*.

Wang, Y., Y. Su, and G. Agrawal. (2013a). "Supporting a Light-Weight Data Management Layer over HDF5". In: *Proceedings of 2013 IEEE/ACM CCGRID International Symposium on Cluster, Cloud and Grid Computing*. 335–342.

Wang, Z., Y. Chu, K. Tan, D. Agrawal, A. E. Abbadi, and X. Xu. (2013b). "Scalable Data Cube Analysis over Big Data". *CoRR*. arXiv:1311.5663.

Whitehorn, M., R. Zare, and M. Pasumansky. (2005). *Fast Track to MDX*. Springer-Verlag.

Widmann, N. and P. Baumann. (1998). "Efficient Execution of Operations in a DBMS for Multidimensional Arrays". In: *Proceedings of 1998 SSDBM International Conference on Scientific and Statistical Database Management*. 155–165.

Wikipedia. (2020). "Basic Linear Algebra Subprograms". URL: https://en.wikipedia.org/wiki/Basic_Linear_Algebra_Subprograms.

Wikipedia. (2022a). "Comparison of OLAP Servers". URL: https://en.wikipedia.org/wiki/Comparison_of_OLAP_servers.

Wikipedia. (2022b). "Math Kernel Library". URL: https://en.wikipedia.org/wiki/Math_Kernel_Library.

Wolf, M. (1989). "More Iteration Space Tiling". In: *Proceedings of 1989 ACM/IEEE SC Conference on Supercomputing*. 655–664.

Wolf, M. E. and M. S. Lam. (1991). "A Data Locality Optimizing Algorithm". In: *Proceedings of 1991 ACM SIGPLAN PLDI Conference on Programming Language Design and Implementation*. 30–44.

Wu, E., S. Madden, and M. Stonebraker. (2013). "SubZero: A Fine-Grained Lineage System for Scientific Databases". In: *Proceedings of 2013 IEEE ICDE International Conference on Data Engineering.*

Wu, K., E. J. Otoo, and A. Shoshani. (2006). "Optimizing Bitmap Indices with Efficient Compression". *ACM Transactions on Database Systems (TODS).* 31(1): 1–38.

Xing, H. and G. Agrawal. (2018). "COMPASS: Compact Array Storage with Value Index". In: *Proceedings of 2018 SSDBM International Conference on Scientific and Statistical Database Management.*

Xing, H. and G. Agrawal. (2019). "Accelerating Array Joining with Integrated Value-Index". In: *Proceedings of 2019 SSDBM International Conference on Scientific and Statistical Database Management.* 145–156.

Xing, H., S. Floratos, S. Blanas, S. Byna, Prabhat, K. Wu, and P. Brown. (2018). "ArrayBridge: Interweaving Declarative Array Processing in SciDB with Imperative HDF5-Based Programs". In: *Proceedings of 2018 IEEE ICDE International Conference on Data Engineering.* 977–988.

Yu, L., Y. Shao, and B. Cui. (2015). "Exploiting Matrix Dependency for Efficient Distributed Matrix Computation". In: *Proceedings of 2015 ACM SIGMOD International Conference on Management of Data.* 93–105.

Yu, Y., M. Tang, W. G. Aref, Q. M. Malluhi, M. M. Abbas, and M. Ouzzani. (2017). "In-Memory Distributed Matrix Computation Processing and Optimization". In: *Proceedings of 2017 IEEE ICDE International Conference on Data Engineering.* 1047–1058.

Yuan, B., D. Jankov, J. Zou, Y. Tang, D. Bourgeois, and C. Jermaine. (2021). "Tensor Relational Algebra for Distributed Machine Learning System Design". *PVLDB.* 14(8): 1338–1350.

Zaharia, M., M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica. (2010). "Spark: Cluster Computing with Working Sets". In: *Proceedings of 2010 USENIX HotCloud Conference on Hot Topics in Cloud Computing.*

Zalipynis, R. A. R. (2018). "ChronosDB: Distributed, File Based, Geospatial Array DBMS". *PVLDB.* 11(10): 1247–1261.

Zalipynis, R. A. R. (2021). “Array DBMS: Past, Present, and (near) Future”. *PVLDB*. 14(12): 3186–3189.

Zee, F. van and R. van de Geijn. (2015). “BLIS: A Framework for Rapidly Instantiating BLAS Functionality”. *ACM Transactions on Mathematical Software (TOMS)*. 41(3).

Zhang, Y., M. Kersten, M. Ivanova, and N. Nes. (2011). “SciQL: Bridging the Gap between Science and Relational DBMS”. In: *Proceedings of 2011 IDEAS Symposium on International Database Engineering and Applications*. 124–133.

Zhao, W., F. Rusu, B. Dong, and K. Wu. (2016). “Similarity Join over Array Data”. In: *Proceedings of 2016 ACM SIGMOD International Conference on Management of Data*. 2007–2022.

Zhao, W., F. Rusu, B. Dong, K. Wu, A. Y. Q. Ho, and P. Nugent. (2018). “Distributed Caching for Processing Raw Arrays”. In: *Proceedings of 2018 SSDBM International Conference on Scientific and Statistical Database Management*.

Zhao, W., F. Rusu, B. Dong, K. Wu, and P. Nugent. (2017). “Incremental View Maintenance over Array Data”. In: *Proceedings of 2017 ACM SIGMOD International Conference on Management of Data*. 139–154.

Zhao, Y., P. M. Deshpande, and J. F. Naughton. (1997). “An Array-Based Algorithm for Simultaneous Multidimensional Aggregates”. In: *Proceedings of 1997 ACM SIGMOD International Conference on Management of Data*. 159–170.